WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

Pooled Testing for HIV Screening: Capturing the Dilution Effect

Lawrence M. Wein
Stefanos A. Zenios

#3665-94-MSA March 1994

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

Pooled Testing for HIV Screening: Capturing the Dilution Effect

Lawrence M. Wein
Stefanos A. Zenios

#3665-94-MSA March 1994

POOLED TESTING FOR HIV SCREENING: CAPTURING THE DILUTION EFFECT

Lawrence M. Wein

Sloan School of Management, M.I.T.

and

Stefanos A. Zenios

Operations Research Center, M.I.T

Abstract

We study pooled (or group) testing as a cost-effective alternative for screening donated blood products (sera) for HIV; rather than test each sample individually, this method combines various samples into a pool, and then tests the pool. A group testing policy specifies an initial pool size, and based on the HIV test result, either releases all samples in the pool for transfusion, discards all samples in the pool, or divides the pool into subpools for further testing. We develop a generalized linear model that relates the HIV test output to the antibody concentration in the pool, and hence captures the effect of pooling together different samples. The model is validated and simplified using data from a variety of field studies, and is embedded into a dynamic programming algorithm that derives a group testing policy to minimize the expected cost due to false negatives, false positives and testing. A simulation study shows that significant cost savings can be achieved without compromising the accuracy of the test. However, the efficacy of group testing depends upon the use of a classification rule (that is, discard the samples in the pool, transfuse them or test them further) that is dependent on pool size, a characteristic that is lacking in currently implemented pooled testing procedures.

February 18, 1994

In the first years of the AIDS epidemic, numerous instances of AIDS infection caused by blood transfusion were reported to the Center for Disease Control. The incidence indicated that the blood supply is a virtually frictionless pathway for spreading the epidemic, and the extent of the epidemic dictated that screening at the individual level should be adopted. As a consequence, all infected blood donors would be identified and a measurably safer blood supply would be attained. Nevertheless, the cost for such a screening program is substantial, and many developing countries, particularly in Africa where the epidemic is spreading rapidly, are struggling to fight the disease on limited budgets.

Pooled testing is one potential way to reduce the monetary cost without compromising the accuracy of the tests. The rationale behind pooled testing is simple and intuitive: suppose we can pool the sera from ten (for example) individuals and test the pool using a single test. If the *seroprevalence* of HIV, which is the fraction of the population that is infected, is low enough, then there is a high probability that all ten individuals in the pool are HIV negative; in this case, we would learn from a single test what otherwise would be learned from ten individual tests. If, on the other hand, the test outcome is positive, then additional tests (either pooled or individual) would need to be carried out.

However, pooled testing has a possible shortcoming, the *dilution effect*: there is a serious concern that if the pool size is too large, then any HIV positive sera will be sufficiently diluted so as to become undetectable by the test. These *false negatives* can be extremely costly, particularly when pooled testing is employed to protect the blood supply. Moreover, infected individuals exhibiting an unusually low level of antibody concentration are less likely to be detected when screened in pools. Consequently, the *sensitivity* of the test can be seriously affected. (*Sensitivity* is the probability of detecting a diseased individual, whereas *specificity* is the probability of detecting a healthy individual.)

Pooling methods have been evaluated in blood banking systems in several developing countries, including Zaire, Zimbabwe and Ecuador (see, for example, Cahoon-Young et al. 1988, Emmanuel et al. 1988, Kline et al. 1989, Behets et al. 1990 and Ledro-Monroy et al. 1990). These field studies suggest that pooling methods (with group sizes as large as 80) may be as sensitive and specific as individual testing, and can result in cost savings from 5% to 80%, depending on the actual seroprevalence. On the other hand, the World Health Organization (WHO), concerned with the dilution effect and its consequences on the sensitivity of the test, is more conservative in their proposal: They recommend the use of pools of size no greater than five in Tamashiro et al. (1993). The discrepancy between the field studies and the recommendations of the WHO is an indication that the dilution effect is not well understood, and the bloodbanking community is in danger of either underestimating or overestimating its importance.

In his seminal paper, Dorfman (1943) showed how pooled testing, which is called *group testing* in the statistical literature, can be employed to efficiently eliminate all defective items from certain large populations. The method found an immediate application in screening World War II draftees for syphilis, where it resulted in considerable savings. The group testing problem was researched aggressively in the 1950's and 60's, (see, for example, Sobel and Groll 1959), and a large literature now exists on this topic; readers are referred to Johnson et al. (1991) for a survey, and to Litvak et al. (1992) for recent work that is motivated by HIV testing. However, nearly all existing studies concentrate on either perfect tests (i.e., tests with no misclassification errors) or imperfect tests with errors that are independent of the group size; two exceptions are Hwang (1976) and Burns and Mauro (1987), who assume that test sensitivity is a specified function of the group size. In addition, all studies neglect the actual test mechanism and, except for Arnold (1977), assume that the test outcome is binary rather than continuous.

In contrast, we attempt to explicitly model both the dilution effect and the continuous nature of the test outcome. Our task is greatly complicated by the fact that the HIV test outcome, which is a continuous quantity called the optical density level, is only an indirect measurement of the unobservable antibody concentration. Starting from first principles, we derive a generalized linear model that relates the optical density level to the antibody concentration of the tested sera. The model explicitly captures the physical pooling of sera, and is validated using data from two existing dilution studies. A simplified version of the model is validated using data from an existing pooling study.

Traditional group testing problems consider a binary test outcome; if the test is HIV negative, then the pool is released for transfusion, and if the test is HIV positive, then the pool is divided into subpools for further testing. If the pool consists of a single sample, then the sample is discarded if the test outcome is HIV positive. Hence, traditional group testing policies are characterized by the initial pool size and the resulting subpool configuration. Because we explicitly consider the continuous nature of the test outcome, our group testing policy must also develop a *classification rule* that is based on the test outcome: the pool is deemed either HIV positive (each sample in the pool is discarded) or HIV negative (each sample in the pool is released for transfusion), or the pool is divided into subpools for further testing. Our validated pooling model is embedded into a dynamic programming framework that derives the group testing policy that minimizes the expected cost due to testing, false positives and false negatives. Our proposed policy is tested on a Monte Carlo simulation model, and the results indicate that pooled testing, with a classification rule that explicitly depends on the pool size, can achieve significant cost savings over individual testing.

The paper is organized as follows. A preliminary description of ELISA, the biological assay used for HIV testing, is given in Section 1. The data used in the paper are described in Section 2. The generalized linear model is developed in Section 3, and is simplified and validated in Section 4 using the data described in Section 2. A dynamic programming framework for the group testing problem is developed in Section 5, and several policies are derived. A simulation study is undertaken in Section 6, and concluding remarks appear in Section 7.

1. Serological Tests for AIDS

The human body reacts to microbial agents, like viruses, bacteria, parasites, etc., by producing antibodies. The antibodies recognize particular molecules on the surface of the infectious agent and bind to them. Such molecules are called *antigens* (*anti*body *gen*erators). Various immunological tests are designed to detect antibodies, thereby identifying the serological status of the individual.

The Human Immunodefficiency Virus (HIV) is the pathological agent of the Acquired Immune Deficiency Syndrome (AIDS). Enzyme Linked ImmunoSorbent Assays (ELISA) for the HIV virus detect the anti-HIV antibodies and are frequently used for HIV screening. This section contains a brief nontechnical description of ELISAs.

A common configuration of ELISAs for HIV is the *indirect assay* pictured in Figure 1 (see George and Schochetman 1985 for more details). Antigens to HIV are attached to a solid phase support (usually wells). The patient's serum (or plasma) is diluted (at a dilution fixed by the manufacturer), added to the solid phase support and incubated for a time period. By the end of the incubation period, any antibodies to HIV that are present in the sample are attached to the antigens on the solid phase support. The well is then washed so that all unattached material is removed, and the attached antibodies become detectable. The attached antibodies (immunoglobulins) are detected when a secondary antibody, labeled by an enzyme, is added. When a substrate is finally added, an enzymatic reaction takes place producing a color change proportional to the amount of human HIV antibodies present. The ELISA test outcome is the *optical density* (OD) level, which quantifies this color change.

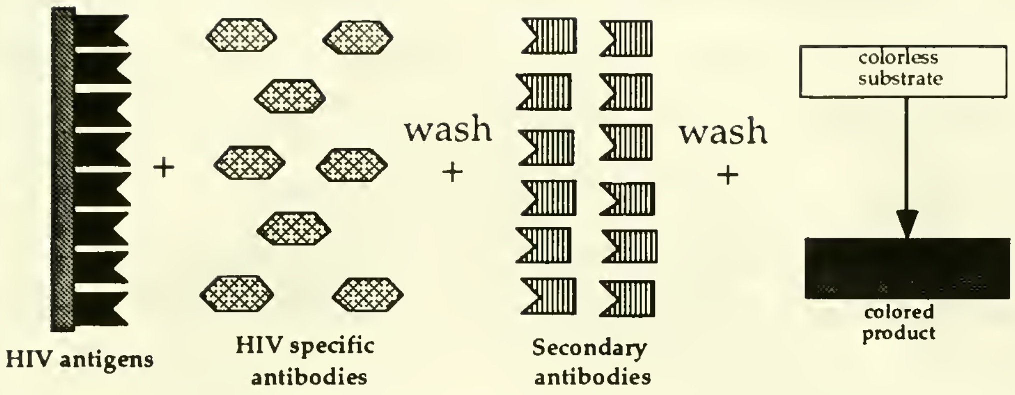


Figure 1: A schematic representation of the indirect ELISA.

Hence, the OD reading is determined by two factors: the concentration of the antibodies and their ability to bind on the antigens on the solid support (antibody affinity). If the OD recorded at the end of the process exceeds the critical value, or cutoff, recommended by the manufacturer, then the patient is declared HIV positive; otherwise the patient is declared HIV negative.

An alternative configuration of ELISAs is the competitive assay. Although the same types of antigens used in indirect ELISAs are attached to the solid phase support, this method differs from the indirect one in the detection mechanism. Enzyme labeled HIV antibodies compete with the patient's antibodies for binding sites. The color change observed is inversely proportional to the concentration of HIV antibodies in the serum. If the recorded OD level exceeds the critical value set by the manufacturer, then the sample is declared negative, otherwise positive. We concentrate on indirect assays in this paper, since most of the commercially available antibody detection kits are based on the indirect configuration of ELISAs. Nevertheless, the study of competitive assays is not any more difficult and only minor modifications, stated when necessary, are required.

ELISAs are inexpensive, easy to administer and very accurate; however, they have a shortcoming that stems from the test's *indirect* detection of HIV via the presence of antibodies. The patient's time of infection is followed by a *window period* during which the antibody concentration in the patient's serum is virtually undetectable. This period usually extends from three to nine months and results in false negatives. Assays for detecting HIV antibodies cannot identify such individuals; therefore, whenever individuals are referred to as positive or negative, we are actually alluding to the presence or absence of HIV antibodies.

2. Description of the Data

We use individual testing data, dilution series data and pooled testing data obtained from three independent sources.

Individual Testing Data. OD readings for 4000 HIV negative and 3000 HIV positive individuals screened using four different assays were provided by the National HIV Reference Laboratory of Australia (Dax 1993). It is convenient to normalize the OD readings according to the equation $x = \frac{OD-A_0}{A_m-A_0}$, so that they fall between zero and one; A_0 and A_m are the minimum and maximum OD readings, respectively, recorded by the assay. The values of A_0 and A_m vary by assay, and were chosen based on an analysis of the data and discussions with the data providers. For this data, we set $A_0 = 0$ and $A_m = 20$.

The empirical distributions for the normalized OD readings for assay A (an indirect ELISA) are given in Figure 2(a). We observe that both the mean and variance are smaller for HIV negative than for HIV positive individuals. The relatively large spread in the HIV positive distribution is to be expected, since an individual's antibody concentration tends to systematically vary as the disease progresses; see George and Schochetman for details. The two populations are well separated, and therefore a critical value separating the OD outcomes into HIV positive and HIV negative can be selected.

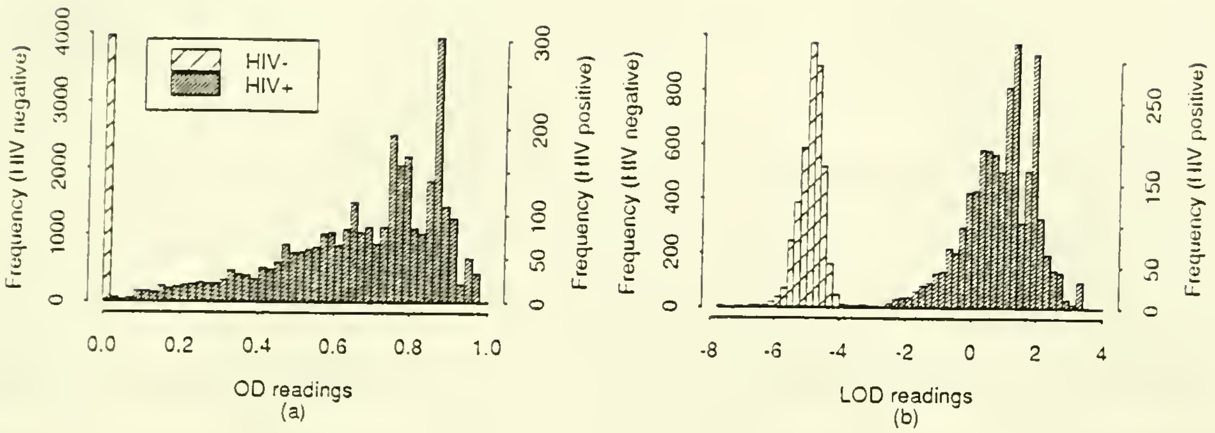


Figure 2: (a) Empirical densities for the reactivity ratios of 4000 HIV negative and 3000 HIV positive individuals, and (b) densities for the corresponding LOD values.

For reasons that will become clear in Section 4, we also consider the *logit* transformation of the normalized OD readings: $x \to \ln(\frac{x}{1-x})$, which will be referred to as the LOD (logit OD) readings. The empirical densities of the LODs for the two populations are given in Figure 2(b). The sample mean and standard deviation are, respectively, $\mu_- = -4.82$ and $\sigma_- = 0.42$ for the HIV negative population, and $\mu_+ = 0.80$ and $\sigma_+ = 1.08$ for the HIV positive population.

Figure 3 displays the normal quantile plot for the empirical distributions in Figure 2(b); that is, the LOD readings are ranked in magnitude and are plotted against the standard normal quantiles. A straight line indicates normality of the data points. The quantile plot of the LOD readings is approximately linear for both populations. Deviations from normality are observed in the tails of the HIV negative population and the right tail of the HIV positive population. Most importantly, the normal approximation captures the left tail of the HIV positive distribution, which contains the low OD readings that might become undetectable under pooled testing. On the other hand, the normal approximation to the HIV

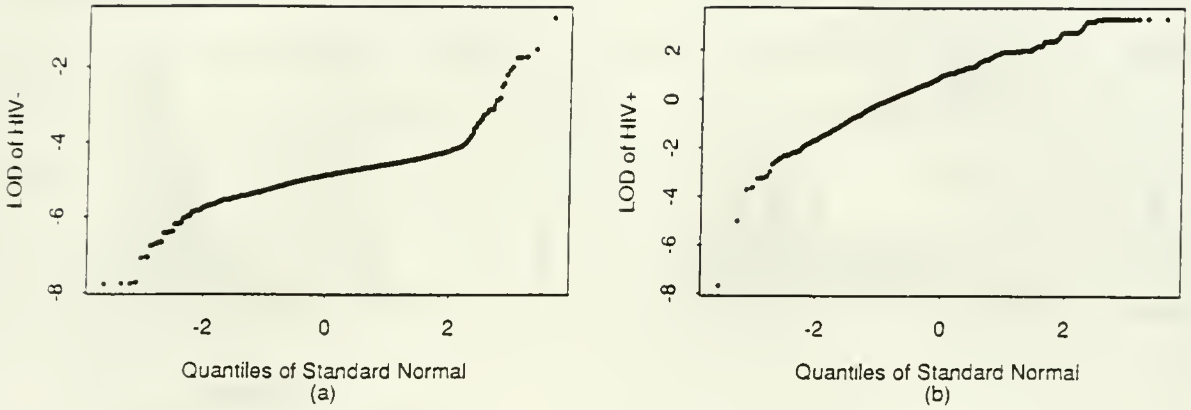


Figure 3: Normal quantiles for the LOD readings of (a) HIV negative and (b) HIV positive populations.

negative population underestimates the proportion of negative individuals with a relatively high OD reading, which might lead to an underestimation of false positives. Nevertheless, the false negatives, which are the overriding concern in pooled testing, will not be affected.

In the analytical model developed in Section 5, we assume that the LOD readings for the HIV negative and positive populations are normally distributed with respective means $\mu_- = -4.82$ and $\mu_+ = 0.80$, and respective standard deviations $\sigma_- = 0.42$ and $\sigma_+ = 1.08$.

Dilution Series Data. Dilution series data were obtained from the Caribbean Epidemiology Center (Hull 1991 and de Gourville 1992) and the National HIV Reference Laboratory of Australia (Dax). The purpose of both of these studies was to investigate the effect of dilution on the ability of ELISAs to detect reactive sera.

In the Caribbean Epidemiology Center (CAREC) study, ten positive sera were diluted sequentially in a fixed negative serum to produce a series of thirteen four-fold dilutions in the ratios $1:1, 1:4, 1:16, \ldots, 1:4^{12}$. A $1:n$ ratio means that $\frac{1}{n}$ of the pool consists of the positive sample. Each dilution was tested by two indirect ELISAs according to the manufacturer's instructions. Since the data from both assays yielded similar results, we only report the results from one of them. The raw data consists of 130 OD readings, one for each of the thirteen dilution levels of each of the ten positive samples. We used $A_0 = 0$ and $A_m = 15$ to normalize the OD readings.

The National HIV Reference Laboratory of Australia (NRL) study sequentially diluted ten positive sera in a fixed negative serum to produce a series of 11 two-fold dilutions, with ratios $1:1, 1:2, 1:4, \ldots, 1:2^{10}$. These dilutions were tested on ten different assays. We analyzed the data from several of these assays and obtained very similar results, and hence will only report on the results from one assay. The OD readings were normalized using $A_0 = 0$ and $A_2 = 2$.

Pooled Testing Data. Cahoon-Young et al. (1992), which will be referred to hereafter as Cahoon-Young et al., tested 1280 specimens individually and in a series of nested pools. More specifically, the individual specimens were pooled to generate 128 pools of size 10; the pools of size 10 were then combined to form 64 pools of size 20, then 32 pools of size 40 and finally 16 pools of size 80. Twelve individuals were found to be HIV positive, and no more than one positive sample was found in any of the pools of size 80. The OD readings at every stage of this nested testing procedure were recorded.

Note that the dilution series studies by CAREC and NRL differ from Cahoon Young et al.'s pooling study in one important respect: positive sera are diluted with varying amounts of the *same* negative sera in the dilution series studies, whereas individual sera are combined with a varying number of *different individual's* sera in Cahoon-Young et al.'s pooling study. Hence, although the two dilution series can be used to assess the effect of dilution, the Cahoon-Young et al. study exactly mimics the pooling that would take place under group testing.

3. A Probabilistic Model for the Dilution Effect

When sera are screened in pools, the OD reading is determined by the concentration and affinity of the antibodies in the pool. The unobservability of the antibody concentration and affinity makes it very difficult to estimate the OD level of a pool. In this section, we develop a stochastic model that predicts the OD level of a sample as a function of the antibody concentration and affinity. We then specialize the model to the setting of the CAREC and NRL dilution studies, and obtain a *generalized linear model* (GLM) that predicts the OD level of a pool as a function of the OD level of the HIV positive sample and the corresponding dilution level. The model is adapted in Section 4 to consider pools consisting of individual samples, as in the Cahoon-Young et al. study.

The dilution series data of CAREC and NRL essentially generate *dose-response curves*: the dose takes the form of a fixed positive sample diluted to various levels, and the response is simply the corresponding OD reading. Empirical dose-response curves typically exhibit sigmoid or hyperbolic behavior, and polynomial, general curvilinear, logit or probit regression models have been proposed to fit these curves. Before our model is introduced, it is worthwhile reviewing the traditional approach, and we focus on the logistic regression for concreteness. Let Y_j denote the LOD reading (that is, the logit of the normalized OD reading, as in Figure 2(b)) of a particular HIV positive sample that is diluted to the ratio $1 : d^j$, where d is an integer ($d = 4$ for CAREC and $d = 2$ for NRL) and $j = 0, 1, \ldots, n$. The linear regression model hypothesizes that

$$Y_j = \alpha + \beta j + \epsilon_j, \tag{1}$$

where ϵ_j are iid normal random variables with zero mean. Although this model typically generates predicted values that coincide well with observed values, it exhibits considerable heteroscedasticity (state-dependent noise), and hence one of the model's basic assumptions is violated (see Tijssen 1985, Chapter 15).

Whereas the existing literature has taken a purely empirical approach to fitting dose-response curves, we develop a probabilistic model that is based upon a set of primitive assumptions regarding the behavior of the ELISA test and the pooled sera. Our analysis leads to a GLM for the dose-response curve: recall that while a linear regression model postulates that the expected value of the dependent variable is a linear function of the independent variable, a GLM assumes that a function of the expected value of the dependent variable is a linear function of the independent variable. Like model (1), our GLM captures the sigmoid nature of the dose-response curve. In addition, it proposes a particular variance function that, as will be seen in the next section, stabilizes the heterogeneous noise present in the CAREC and NRL data sets.

Our model is influenced by Fisher (1922), who developed a probabilistic model to estimate the number of bacteria in a sample of water or soil. The following eight assumptions are used to derive our model. We conferred with several specialists, and none of these assumptions generated any disputation; assumption 5 was the only one that appeared to stimulate any reflection.

A1. The number of HIV antigens, n, attached to a well satisfies $n > 10^6$.

A2. No more than one HIV antibody can bind to any antigen.

A3. The probability of an antibody binding to a specific antigen is small, independent for each antigen, and constant for antibodies from the same serum. The expected number of attached antibodies on a large collection of antigens is significant.

A4. A secondary antibody will bind to *all* attached primary antibodies.

A5. The normalized OD reading is linearly proportional to the number of attached antibodies.

A6. The expected number of attached HIV antibodies is linearly proportional to the antibody concentration. The proportionality constant can vary among different individuals due to differences in their binding properties (affinity).

A7. The antibody concentration of pooled sera is the weighted average of the individual antibody concentrations.

A8. Measurement errors are negligible.

If a competitive assay is employed, then assumption *A5* is replaced by

A5c. The normalized OD reading is proportional to the number of attached secondary antibodies,

and the following assumption is introduced:

A9. The affinity of primary antibodies is higher than that of secondary antibodies; therefore, secondary antibodies will bind to antigens on the solid support if and only if there are not enough primary antibodies to saturate the binding sites.

In this case, the subsequent model derivation follows virtually unchanged.

Motivated by Fisher's analysis, we consider a well with n antigens bound on it, and introduce a *partition* of the well into k subwells indexed $s = 1, \ldots, k$. Assuming that the antigens are uniformly bound on the well, we let $m = \frac{n}{k}$ denote the number of antigens on every subwell. Suppose that serum i is diluted with an HIV negative serum in the ratio $1 : d^j$, and then added to the well. Since neither the concentration nor the affinity of antibodies is observable, and since their net effect is multiplicative in nature by *A6*, we will hereafter refer to the product of the antibody concentration and the antibody affinity as the antibody concentration. Let ρ_{io} denote the antibody concentration for the undiluted serum i, $\rho_{i\infty}$ be the antibody concentration of the HIV negative serum, ρ_{ij} be the antibody concentration for the diluted serum, and p_{ij} be the binding probability for the antibodies in the serum; note that none of these quantities are observable. Also, let N_{ijs} represent the number of antibodies attached to the antigens on subwell s.

Our model development consists of three main steps: First, we find the probability distribution for $S_{ijk} = (N_{ij1} + \ldots + N_{ijk})/k$, which is the average number of attached antibodies per subwell, in terms of the unknown binding probability p_{ij}. Then we use assumption *A6* to relate the unknown binding probability p_{ij} to the unknown antibody concentration in the diluted serum. Finally, we use assumption *A5* to relate the average number of attached antibodies per subwell S_{ijk} to the normalized OD reading. Combining these three steps yields our basic model.

By our comments above, $N_{ij1}, \ldots, N_{ijk}$ are independent binomial random variables with size parameter m and success probability p_{ij}. By *A3* and the law of rare events, the binomial random variable can be approximated by a Poisson random variable:

$$P(N_{ijs} = k) \approx e^{-p_{ijm}} \frac{p_{ijm}^k}{k!}, \tag{2}$$

where $p_{ijm} = mp_{ij}$. We can choose a sufficiently fine partition of the well such that $P(N_{ijs} > 1) = o(p_{ijm})$, implying

$$\begin{aligned} P(N_{ijs} = 0) &\approx e^{-p_{ijm}}, &(3)\\ P(N_{ijs} = 1) &\approx 1 - e^{-p_{ijm}} \text{ and} &(4)\\ P(N_{ijs} > 1) &\approx o(p_{ijm}). &(5) \end{aligned}$$

By the Central Limit Theorem and (3)-(5),

$$S_{ijk} \overset{law}{\to} N(1 - e^{-p_{ijm}}, \frac{1}{k} e^{-p_{ijm}}(1 - e^{-p_{ijm}})) \text{ as } k \to \infty, \tag{6}$$

and hence

$$\ln(E(1 - S_{ijk})) = -p_{ijm}. \tag{7}$$

Since $\sum_{s=1}^{k} N_{ijs}$ is a binomial random variable, the distribution of S_{ijk} is well approximated by the normal distribution even for relatively small values of k (typically $k \geq 15$). In Section 4, the parameters of our resulting model are estimated from the data, and k approximately equals 20.

Assumption *A6* implies that

$$p_{ijm} = \frac{\rho_{ij}}{k}, \tag{8}$$

which relates the binding probability to the antibody concentration in the diluted serum.

Let X_{ij} denote the normalized OD reading of the diluted sample. Then $A5$ implies that $X_{ij} = \gamma S_{ijk}$, where γ is the constant of proportionality. The maximum normalized OD reading, $X_{ij} = 1$, is attained when antibodies are bound on all subwells. In this case, $S_{ijk} = 1$ by (3)-(5), and so $\gamma = 1$ and

$$X_{ij} = S_{ijk}. \tag{9}$$

Combining equations (6) and (8)-(9) and taking logarithms gives the basic stochastic model relating the normalized OD reading to the antibody concentration:

$$X_{ij} \sim N(E[X_{ij}], \frac{1}{k}E[X_{ij}](1 - E[X_{ij}])), \tag{10}$$

where

$$\ln\Big(-\ln(1 - E[X_{ij}])\Big) = \ln\left(\frac{\rho_{ij}}{k}\right). \tag{11}$$

In Section 4, this basic model will be adapted to the Cahoon-Young et al. pooling study. Now we specialize this model to the setting of the CAREC and NRL dilution studies. By $A7$, the antibody concentration in the diluted serum is

$$\begin{aligned} \rho_{ij} &= \frac{\rho_{i0} + (d^j - 1)\rho_{i\infty}}{d_j} \\ &\approx \frac{\rho_{i0}}{d^j}(1 + d^j \frac{\rho_{i\infty}}{\rho_{i0}}). \end{aligned} \tag{12}$$

Combining equations (11) and (12), and using the first order Taylor series approximation $\ln(1+x) \approx 0$ (the adequacy of this approximation will be investigated during the model validation phase) gives the GLM

$$\ln\Bigl(-\ln(1-E[X_{ij}])\Bigr) = \ln\left(\frac{\rho_{i0}}{k}\right) - j\ln d. \tag{13}$$

The random component of the model is the normalized OD level X_{ij} of the diluted sample, which is normally distributed by (10). The systematic component is the dilution level j. The link between the random component and the systematic component is of the form

$$g(\mu_{ij}) = \alpha_i + \beta j \tag{14}$$

where $\mu_{ij} = E(X_{ij})$, the link function is $g : x \rightarrow \ln(-\ln(1-x))$, the complementary log-log (cloglog), and $\alpha_i = \ln(\frac{\rho_{i0}}{k})$ and $\beta = -\ln d$ are constants. The second moment of the random component is given by $Var(X_{ij}) = \phi\mu_{ij}(1-\mu_{ij})$, which will be denoted by $V(\mu_{ij})$, where the *dispersion parameter* ϕ equals $\frac{1}{k}$.

We conclude this section with several remarks about the GLM (13). It captures the sigmoid nature of the dose-response curve via the cloglog link function. Other suitable sigmoid link functions are the *logit* and *probit*. The logit link is defined by $g : x \rightarrow \ln(\frac{x}{1-x})$ and the probit by $g : x \rightarrow \Phi^{-1}(x)$, where Φ is the cumulative standard normal distribution. To obtain the best fit of the data, GLMs for all three link functions will be considered in Section 4. Notice that if we replace the cloglog function by the logit function in (11) and set the dilution level $j = 0$, then this equation implies that individual LOD readings are normally distributed, which is consistent with the assumption that was discussed in Section 2, and that will be used in Section 5.

The y-axis intercept α_i corresponds to the cloglog of the normalized OD level of the positive sample, and hence provides a measure of the antibody concentration of the original positive sample; our model predicts that $\alpha_i = \ln(\rho_{i0}/k)$, which is perfectly consistent with this interpretation. The slope β is the marginal change in response due to a change in the dilution level j; hence, $\beta = -\ln d$ is also consistent with this interpretation. Notice that ρ_{i0} and k are not observable, and d is observable. Hence, the slope of the model is known, but the $y-$intercept and the dispersion parameter are unknown. In the next section, we fit the GLM to the CAREC and NRL data sets. Unfortunately, it is very tedious to estimate α_i and ϕ for a fixed slope $\beta = -\ln d$. Therefore, we will estimate α, β and ϕ from the model, and see if the predicted slope is close in value to $-\ln d$.

4. Model Validation

In this section, we attempt to validate the GLM developed in Section 3. In Subsection 4.1, the parameters of the model are estimated using the dilution series studies by CAREC and NRL. The GLM is adapted to the pooling setting and simplified in Subsection 4.2. The simplified pooling model is validated on Cahoon-Young et al.'s data in Subsection 4.3.

4.1 Model Fitting

The dilution series studies undertaken by CAREC and NRL generate normalized OD values X_{ij} for positive sample i and dilution level j. This data will be used to estimate the values of the generalized linear model parameters α, β and ϕ. If the random mechanism embodied in the GLM is the true process by which the data are generated, then the maximum likelihood estimators can be obtained by iterative, weighted least squares. However, the normality assumption on X_{ij} is approximate, and we can relax this assumption by employing the theory of *quasilikelihood functions* (see pp. 323-352 of McCullagh and Nelder 1989).

This theory applies under the following four conditions that are satisfied by the GLM: (i) the range of possible normalized OD values X_{ij} is known, (ii) the mean normalized OD level is specified as a function of the dilution level j, (iii) the variance of the normalized

OD is specified as a function of the mean OD, and (iv) the observations are statistically independent. Let us fix sample i, and let $X_i = (X_{i0}, X_{i1}, \ldots, X_{in})$ denote the random vector of normalized OD readings, and assume that the X_{ij}'s are independent with mean μ_{ij} and variance $V(\mu_{ij})$. Then the log-likelihood function for μ_{ij} can be replaced by the quasilikelihood function

$$\int_{x_{ij}}^{\mu_{ij}} \frac{x_{ij} - y}{V(y)} dy, \tag{15}$$

where x_{ij} is the realization of X_{ij}. The maximum likelihood estimators (MLE) for the model parameters are then obtained by maximizing the quasilikelihood function

$$Q(\mu, x) = \sum_{i=1}^{10} \sum_{j=1}^{n} \int_{x_{ij}}^{\mu_{ij}} \frac{x_{ij} - y}{V(y)} dy \tag{16}$$

for each study, where $n = 12$ for CAREC and $n = 10$ for NRL. We use the *glm* routine of S-plus (see Hastie and Pregibon 1992) to obtain MLEs for the dispersion parameter ϕ, the slope β and the $y-$intercept $\alpha_i, i = 1, \ldots, 10$ for each study. The *predicted values* $\hat{\mu}_{ij}$ are the mean response values predicted by the model, and are given by

$$\ln\left(\frac{\hat{\mu}_{ij}}{1 - \hat{\mu}_{ij}}\right) = \hat{\alpha}_i + \hat{\beta} j. \tag{17}$$

The Pearson residuals are defined as $\frac{x_{ij} - \hat{\mu}_{ij}}{V(\hat{\mu}_{ij})}$.

Since most GLM diagnostics are visual, we begin by analyzing the scatter plot of the predicted vs. observed values and the Pearson residual plots. Our analysis is illustrated using the logit link function. The scatter and residual plots for CAREC and NRL are given in Figure 4. No significant deviations from the predicted fit (the dotted line) are observed in the scatter plots. The observations are fairly uniformly spread along the fitted line and no outliers are detected. It is worth noting that the traditional linear regression model (1) was also fit to the data, and severe heteroscedasticity was present; hence, the variance function $V(\mu)$ appears to stabilize the residuals, giving an almost uniform spread of the residuals around zero. Moreover, as illustrated in Figure 5, the three link functions under consideration are

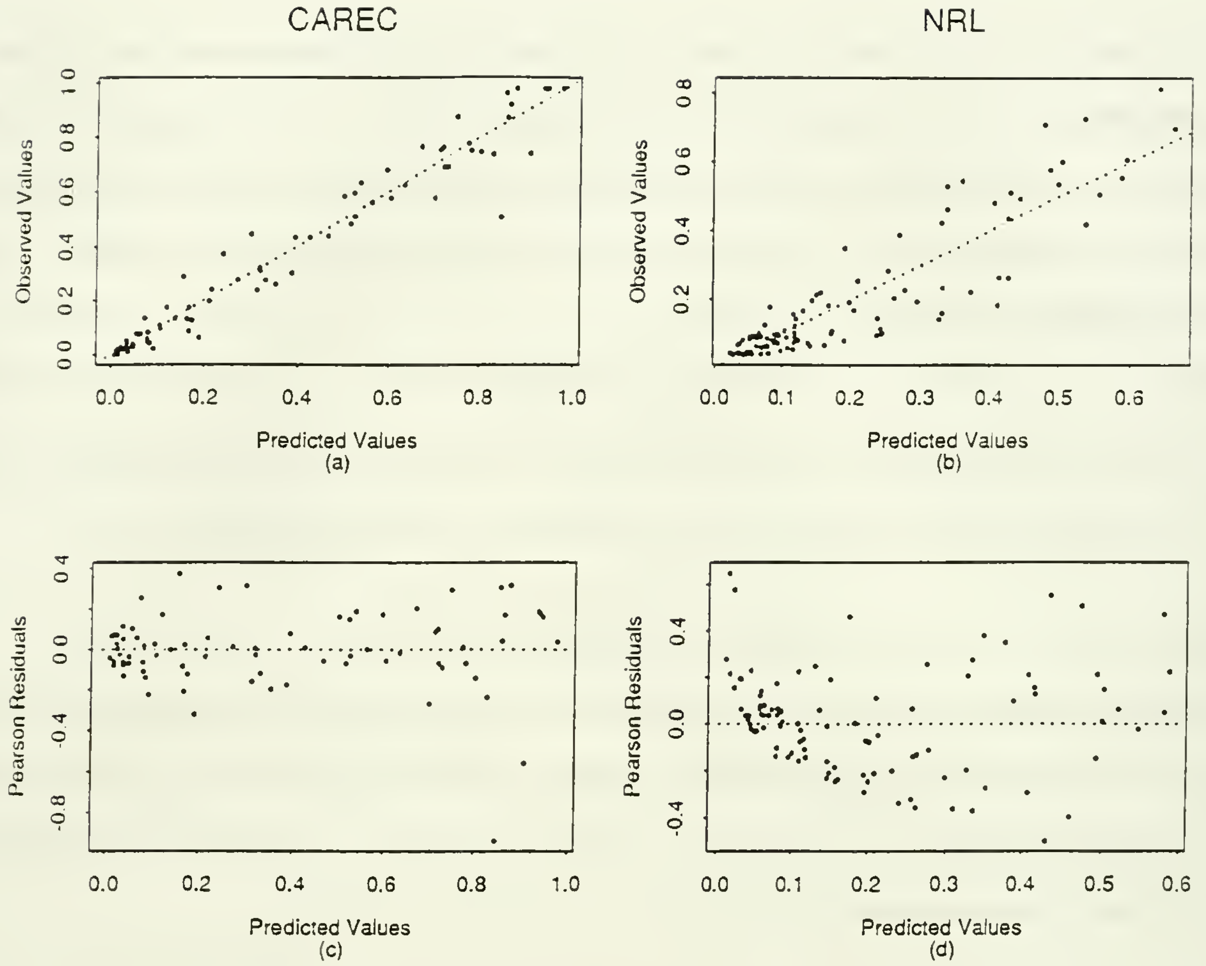


Figure 4: Residual and scatter plots for CAREC (a,c) and NRL (b,d).

nearly equivalent, in the sense that the resulting models predict very similar values; data in Figure 5 are from the NRL study, which exhibited less agreement among the link functions than the CAREC study. Hereafter, our analysis will employ the logit link function. In summary, the GLM provides a reasonably good fit to the available data, particularly for the CAREC data set, and several diagnostics did not reveal any serious model inconsistencies.

To assess the predictive power of the model, we use the LOD readings at dilution levels $j = 1, 2, 6, 7, 11, 12$ from CAREC and $j = 1, 2, 5, 7$ from NRL to obtain the MLEs for the logit model. The predicted values obtained for the remaining data (i.e., dilution levels $j = 3, 4, 5, 8, 9, 10$ for CAREC and $j = 3, 4, 6, 8, 9, 10$ for NRL) are plotted against the

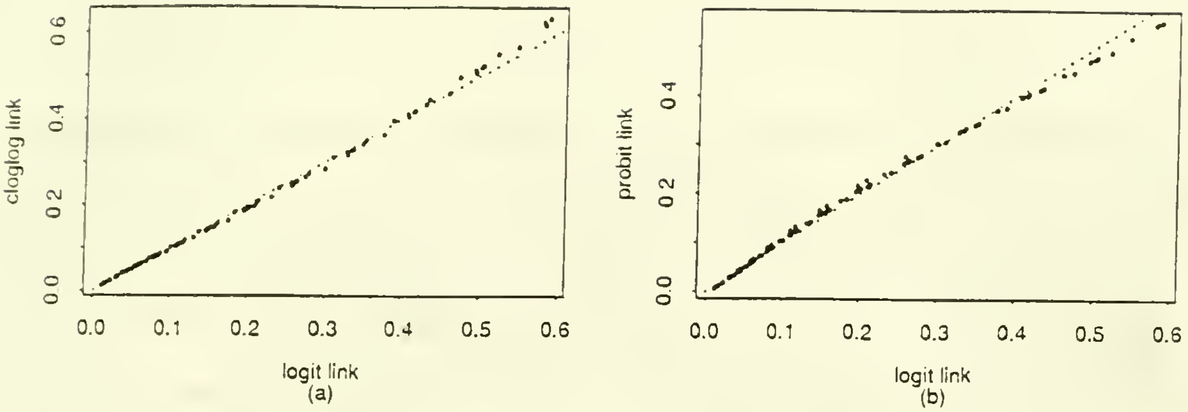
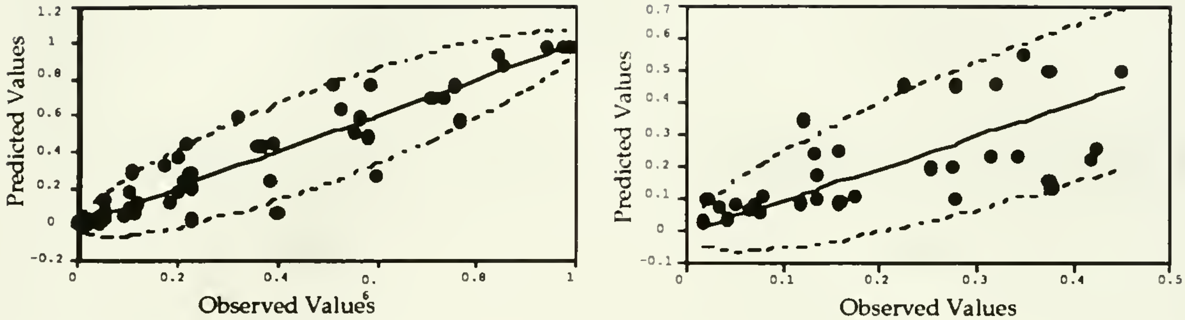


Figure 5: Scatter plots for the response predicted by the logit, cloglog and probit link functions.

observed values in Figure 6. The predictive power of the model is verified by observing that nearly all the observed points lie within the 95% confidence interval predicted by the model.

The visual diagnostics can be supplemented by quantitative diagnostics based on the residual deviance, which is twice the log-likelihood ratio. In particular, the goodness-of-fit can be assessed quantitatively by comparing the proposed GLM to the *full model* that has as many parameters as observations. The residual deviance is asymptotically χ^2 with degrees of freedom given by the difference between the number of observations and the number of model parameters. Table 1 shows the 95% confidence intervals for β, the MLE for the dispersion parameter ϕ, the degrees of freedom of the χ^2 statistic and the residual deviance. The quality of the best fit is ascertained by the significance level of the χ^2 statistic, which is greater than 0.999 in all cases.

Recall that equation (13) predicts that $\beta = -\ln d$. Since the 95% confidence intervals in Table 1 do not contain $-\ln 4 = -1.386$ for CAREC and $-\ln 2 = -0.693$ for NRL, we deduce that $-\ln d$ is not an accurate prediction of the slope that best fits the GLM. However, by deriving an upper bound on the deviance of the fixed slope GLM, we can show that even



(a) CAREC

(b) NRL

Figure 6: Verifying the predictive power of the GLM: The observed values lie within the 95% confidence interval predicted by the model.

for $\beta = -\ln d$, the GLM accurately describes the data. The bound is obtained by replacing the actual MLEs by a set of readily available suboptimal estimates.

	CAREC				NRL			
Model	95% CI for β	$\hat{\phi}$	d.f.	χ^2	95% CI for β	$\hat{\phi}$	d.f.	χ^2
cloglog	-0.6020 ± 0.065	0.065	68	3.2130	-0.3081 ± 0.042	0.052	88	4.4103
logit	-0.8323 ± 0.078	0.042	68	2.6393	-0.3436 ± 0.052	0.056	88	4.4758
probit	-0.4883 ± 0.041	0.042	68	2.6802	-0.1880 ± 0.032	0.065	88	5.3035

Table 1: GLM parameter estimates and goodness of fit.

Let $\hat{\alpha}_i$ and $\hat{\beta}$ be the MLEs for the variable slope GLM (14), and let α_i^* be the MLEs for the GLM that has the slope fixed at $\beta = -\ln d$. The deviance for the fixed slope GLM is $Q_0 - Q(\mu^*, x)$, where Q_0 is the maximum quasilikelihood for the full model and μ^* is the set of mean response values predicted using the MLEs α_i^* and $\beta = -\ln d$. Suboptimal estimates for the MLEs of the fixed slope model can be derived by using $\alpha = \hat{\alpha}$ and $\beta = -\ln d$. Let $\tilde{\mu}$

be the set of mean response values predicted using the suboptimal estimates; then $Q(\tilde{\mu}, x)$, which is the quasilikelihood for the suboptimal estimates, is bounded below by $Q(\mu^*, x)$. Therefore, $Q_0 - Q(\tilde{\mu}, x)$ is an upper bound on the deviance of the fixed slope GLM. The upper bound for the logit model is 46.2412 for CAREC and 13.617 for NRL. Since the corresponding significance levels are 0.9996 and 1.0 respectively, we deduce that the fixed slope model provides a reasonable description of the data.

The quasilikelihood analysis can also provide useful insights into the mechanism of ELISAs and the reliability of the test outcomes. Consider the coefficient of variation of the normalized OD level X_{ij}, which is

$$c_v = \sqrt{\frac{\phi(1 - \mu_{ij})}{\mu_{ij}}}. \tag{18}$$

Equation (18), the values of $\hat{\phi}$ in Table 1, and Figure 4 suggest that the coefficient of variation is quite small for HIV positive samples that have not been substantially diluted, whereas it is much larger (near one) for HIV negative samples or highly diluted positive samples. This observation will be instrumental in the development of the Monte Carlo simulation model in Subsection 6.2. As a side remark, since $\phi = \frac{1}{k}$, the value of k is at least 15, and hence the Central Limit Theorem employed in (6) is a reasonable approximation.

Two variants of the model were also considered in a failed attempt to obtain a more accurate fit. Recall that (13) was derived under the rather crude approximation $\ln(1+x) \approx 0$. For high dilution levels, the assumption $\rho_{i0} \gg \rho_{i\infty} d^j$ (see equation (12)) underlying this approximation will be violated. Employing the second order Taylor approximation $\ln(1+x) \approx x$, we tested the refined model

$$\ln\Big(-\ln(1 - E[X_{ij}])\Big) = \ln(\frac{\rho_{i0}}{k}) - j \ln d + d^j \frac{\rho_{i\infty}}{\rho_{i0}}. \tag{19}$$

Residual plots and scatter plots that are not displayed here indicate that for all three link functions, this refinement has very little effect on the quality of the model fit. We also tested the alternative variance function $V(\mu) = \phi\mu^2(1-\mu)^2$ on the GLM with all three link functions. The deviance for the variance function $V(\mu) = \phi\mu(1-\mu)$ is significantly smaller than the corresponding deviance for $V(\mu) = \varphi\mu^2(1-\mu)^2$, indicating that the original variance function provides a better description of the data.

4.2. A Simplified Pooling Model

The complexity of the traditional (i.e., where no dilution effect is captured) group testing problem, combined with the complexity of the GLM, leads to an analytically intractable model for analyzing group testing policies for ELISAs. Consequently, we propose two simplifications of the GLM that will allow for a tractable analysis in Section 5. This simplified model will be validated on the Cahoon-Young et al. data in Subsection 4.3.

The first simplification is rather bold: Motivated by our earlier observation that the dispersion estimates $\hat{\phi}$ in Table 1 are small, we propose to ignore the variability in the GLM and employ a deterministic model that provides a one-to-one mapping between normalized OD readings and antibody concentrations. Although this assumption compromises the accuracy of the GLM for the sake of tractability, the discussion about the coefficient of variation c_v below equation (18) suggests that the resulting deterministic model should be reasonably reliable for the dilution of HIV positive samples at practical dilution levels. Our assumption leads to the simplified dilution model

$$\ln\left(\frac{X_{ij}}{1-X_{ij}}\right) = \ln\left(\frac{\rho_{i0}}{k}\right) - j\ln d, \tag{20}$$

where the logit, rather than the cloglog, function is being employed.

Recall that this dilution model is appropriate for the CAREC and NRL dilution series, where a given positive serum is diluted with a varying amount of a fixed HIV negative serum, but is not appropriate for Cahoon-Young et al.'s data, which mirrors an actual pooled test. We now present a variant of our model that is appropriate for pooled testing. Let a pool consist of n samples that have individual normalized OD readings $X_1, \ldots, X_n$ and antibody concentrations $\rho_1, \ldots, \rho_n$. Let X and ρ denote the normalized OD level and the antibody concentration, respectively, of the pool. The only difference between the dilution GLM (13) and the pooling GLM is that equation (12) is replaced by $\rho = (\rho_1 + \ldots + \rho_n)/n$. Repeating the steps leading from (11) to (13) gives the pooling GLM

$$\ln\left(\frac{E[X]}{1 - E[X]}\right) = \ln\left(\frac{\rho_1 + \ldots + \rho_n}{n}\right) - \ln k, \tag{21}$$

and ignoring the variability in this GLM leads to

$$\ln\left(\frac{X}{1 - X}\right) = \ln\left(\frac{\rho_1 + \ldots + \rho_n}{n}\right) - \ln k, \tag{22}$$

which is the pooling analog to the simplified dilution model (20).

Our second simplification is to replace the logarithm of the average antibody concentration in (22) by a linear approximation, which yields

$$\ln\Big(\frac{X^p}{1 - X^p}\Big) = \frac{1}{n}(\ln \rho_1 + \ldots + \ln \rho_n) - \ln k. \tag{23}$$

Since $\ln \rho_i = \ln(\frac{X_i}{1-X_i}) + \ln k$ for $i = 1, \ldots, n$ by (22), we obtain the *simplified pooling model*

$$\ln\left(\frac{X^p}{1 - X^p}\right) = \frac{1}{n}\Big(\ln(\frac{X_i}{1 - X_i}) + \ldots + \ln(\frac{X_i}{1 - X_i})\Big). \tag{24}$$

The concavity of the logarithmic function implies that $\ln(\frac{\rho_1 + \ldots + \rho_n}{n}) \geq \frac{(\ln \rho_1 + \ldots + \ln \rho_n)}{n}$; hence, the linearity assumption is conservative in that it underestimates the OD for the pooled sera, and provides an upper bound on the number of false negatives that result as a consequence of pooled testing. Our simplified pooling model (24) has an interesting and tractable characterization: the LOD (logit of the normalized OD) reading for a pool is given by the average of the individual LOD readings.

4.3. Validation of the Simplified Pooling Model

The simplified pooling model (24) will be validated on the Cahoon-Young et al. data in this subsection. The randomness in the individual LOD readings, when embedded into the deterministic pooling model, leads to a random walk model for the pooling data. We fit the random walk model to the pooling data, and use a nonparametric approach to test whether the increments of the random walk are independent and have zero mean.

Recall that Cahoon-Young et al. individually tested 1280 samples, and then generated nested pools of size 10, 20, 40 and 80. The total sample contained 12 HIV positive individuals, and none of the pools contained more than one positive sample; hence, 12 of the 16 pools of size 80 contained exactly one positive sample.

We only test the random walk model on the pools that contain one positive sample. In fact, the Cahoon-Young et al. data is incomplete: Two of the 12 positive samples were not tested at all pool sizes. Hence, we will restrict ourselves to the ten samples that were tested at all pool sizes. Let $i = 1, \ldots, 10$ index the ten positive samples. Let Y_{i1} denote the LOD reading for positive sample i, Y_{ij} be the LOD reading of the j^{th} negative sample pooled with positive sample i, and Y^p_{is} denote the LOD reading for the pool of size 10×2^s containing positive sample i; this pool consists of samples $Y_{i1}, \ldots, Y_{i,10\times 2^s}$, where $s = 0, 1, 2, 3$.

Notice that for fixed i, Y_{ij} are iid random variables for $j = 2, 3, \ldots, 80$, since they all correspond to LOD readings for negative sera; the assumption of normal LOD readings is not required in this subsection. Let $\mu = E(Y_{ij})$ for $j > 1$. For $i = 1, \ldots, 10$ and $s = 1, 2, 3$, the simplified pooling model (24) implies that

$$Y^p_{is} = \frac{\sum_{j=1}^{10\times 2^s} Y_{ij}}{10 \times 2^s} \tag{25}$$

$$= \frac{1}{2}\frac{\sum_{j=1}^{10\times 2^{s-1}} Y_{ij}}{10 \times 2^{s-1}} + \frac{\sum_{j=10\times 2^{s-1}+1}^{10\times 2^s} Y_{ij}}{10 \times 2^s} \tag{26}$$

$$= \frac{1}{2} Y^p_{i,s-1} + \epsilon_{is}, \tag{27}$$

where $\epsilon_{is} = \frac{\sum_{j=10\times 2^{s-1}+1}^{10\times 2^s} Y_{ij}}{10\times 2^s}$ is the *noise* term. Since $E(\epsilon_{is}) = \frac{1}{2}\mu$, equation (27) describes a three-step random walk $(Y_{i0}^p, Y_{i1}^p, Y_{i2}^p, Y_{i3}^p)$ that starts at Y_{i0}^p, ends at Y_{i3}^p and has drift $\frac{1}{2}\mu$. The following proposition shows that the autoregressive process (27) can be transformed into an equivalent driftless random walk by defining $V_{is} = 2^s(Y_{is}^p - \mu)$.

Proposition 1 *For $i = 1, \ldots, 10$, $(V_{i0}, V_{i1}, V_{i2}, V_{i3})$ is a three-step driftless random walk.*

The proof can be found in the Appendix. The simplified pooling model (24) can be validated by establishing that the driftless random walk model describes the Cahoon-Young et al. pooling data. This validation is accomplished by verifying that the random increments are independent and have zero mean. Notice that the random walk only models the pooling effect as the pool size changes from 10 to 20, 20 to 40, and 40 to 80; it does not capture the pooling effect when the pool size changes from 1 to 10.

Hypothesis I: Independent Random Increments. A point estimate for μ is required to pursue a statistical analysis. The following proposition suggests that the estimator should be chosen to minimize the variance.

Proposition 2 *Define $V_{is}(x) = 2^s(Y_{is}^p - x)$; then $\mu = arg\ min_x E(V_{is}(x) - V_{i0}(x))^2$.*

Since the true variance is not available, we consider the estimator that minimizes the sample variance. An additional degree of freedom can be introduced by considering the weighted sum of squares

$$\sum_{i=1}^{10}\sum_{s=1}^{3} w_s(V_{is}(x) - V_{i0}(x))^2. \tag{28}$$

The weights w_1, w_2, w_3 can be chosen in such a way that the sample variance of the estimator is minimized. The weighted least squares estimator $\hat{\mu}$ minimizing (28) is given by

$$\hat{\mu} = -\frac{\sum_{i=1}^{10}\sum_{s=1}^{3}[w_s\{2^s(1-2^s)Y_{is}^p - (1-2^s)Y_{i0}^p\}]}{10\sum_{s=1}^{3}\{w_s(1-2^s)^2\}}, \tag{29}$$

and the following two propositions characterize its statistical properties.

Proposition 3 *The estimator $\hat{\mu}$ is an unbiased estimator of μ for any choice of weights w_i.*

Proposition 4 *The most efficient estimator for the pooling data is obtained for weights $w_1 = w_2 = 0, w_3 = 1$.*

The proofs can be found in the Appendix.

Let us now consider the process $\hat{V}_{is} = 2^s(Y_{is}^p - \hat{\mu})$. As the number of positive samples in the data tends to infinity, the Strong Law of Large Numbers implies that $\hat{\mu} \rightarrow \mu$ (and hence $\hat{V}_{is} \rightarrow V_{is}$) with probability one. Although we have only ten positive samples, we will carry out a statistical hypothesis test on $\hat{V}_{is}$ and extrapolate the conclusions to V_{is}.

The independence assumption of the random walk will be tested by studying the increments $\Delta\hat{V}_{is} = \hat{V}_{is} - \hat{V}_{i,s-1}$ for $s = 1, 2, 3$. If the increments are independent, then with probability $\frac{1}{2}$ they are either above or below the median. We use the *runs above or below the median* test, which is a nonparametric procedure that is described in detail in Chapter 4 of Madansky (1988). The test is applied as follows. Let u_{is} take the value 1 if $\Delta\hat{V}_{is}$ is above the sample median and 0 otherwise. A *run* is a maximal consecutive set of random components u_{is} having the same value. For example, if the sequence is 1001001110111100110100, then the runs are separated as follows: 1|00|1|00|111|0|1111|00|11|0|1|00, and the run count is 12. A low run count is an indication that observations below or above the median come together, whereas a high count is typical for median reverting behavior. Note that under the independence hypothesis, the run count for the random vector $(\Delta\hat{V}_{i1}, \Delta\hat{V}_{i2}, \Delta\hat{V}_{i3})$ is k with probability p_k, where $p_1 = p_3 = \frac{1}{4}$ and $p_2 = \frac{1}{2}$. Define T_k to be the number of random vectors in our data set with run count k, where $\sum_{k=1}^{3} T_k = 10$. Then

$$P(T_1, T_2, T_3) = \frac{10!}{T_1!T_2!T_3!} p_1^{T1} p_2^{T_2} p_3^{T_3} \tag{30}$$

is the significance of the observations under the *null hypothesis*.

For our data, we calculated $T_1 = 2, T_2 = 5, T_3 = 3$; the significance of these observations is $P(T_1 = 2, T_2 = 5, T_3 = 3) = 0.077$. Although a $p-$value for this test cannot be obtained without ordering the state space, the probability of observing an outcome as extreme as this under the independence assumption is at least 0.077; hence, we cannot reject the independence assumption at the 95% significance level. In fact, the outcome $(2, 5, 3)$ is, along with outcomes (3,5,2) and (2,6,2), the mode of the distribution of (T_1, T_2, T_3).

Hypothesis II: Zero mean random increments. The 95% confidence intervals for the mean of $\Delta\hat{V}_{i1}$, $\Delta\hat{V}_{i2}$ and $\Delta\hat{V}_{i3}$ are given by 0.2065 ± 0.2529, 0.08032 ± 0.4961 and -0.3491 ± 1.0637, respectively. The zero mean hypothesis cannot be rejected, since zero is a common point of the three intervals. In conclusion, the data support the hypothesis that the random walk (27) provides a realistic description, thus establishing their consistency with the simplified pooling model (24).

5. The Derivation of Group Testing Policies

In this section, we embed the simplified pooling model (24) into an optimization framework to find efficient pooled testing policies. Our objective is to minimize the expected weighted cost due to testing, false positives and false negatives. Suppose a pool of a specified size is tested and its LOD reading is determined. The decision maker has three options: stop testing and classify all individuals in the pool as HIV negative (and transfuse these samples), stop testing and classify all individuals in the pool as HIV positive (and discard these samples) or divide the pool into subpools for further testing. There are many possible ways to subdivide the pool under the third option, and we consider a quite general class of multistage policies employed by Arnold, where each subpool is of identical size. Our procedure can be modified slightly to allow unequal subpool sizes, as in the sequential procedure in Hwang (1984) and the $T_s^r(V)$ procedure in Litvak et al.

For a given initial pool size and subpool configuration, a dynamic programming algorithm is developed in Subsection 5.1 for finding the optimal policy within the class of multistage policies under consideration. Exhaustive search among alternative initial pool sizes and subpool configurations is required to find the cost minimizing policy. Structural properties of the optimal policy are investigated in Subsection 5.2. Since the dynamic programming algorithm is computationally intensive and the resulting policy difficult to implement, a procedure for deriving near optimal Dorfman policies is derived in Subsection 5.3.

5.1. The Dynamic Programming Formulation

We assume that the blood donor population is composed of two subpopulations: HIV negative (denoted P_-) and HIV positive (P_+). The LOD readings of P_- (P_+, respectively) are assumed to be iid normal random variables with mean μ_- (μ_+, respectively) and standard deviation σ_- (σ_+, respectively). This assumption is consistent with the GLM and provides a reasonable fit to the data in Figure 2. The known *seroprevalence* π is the probability that a random donor is HIV positive. Arnold's notation will be adopted to describe the multistage testing procedure. Consider a random sample of $n_1 = \prod_{j=1}^{N} a_j$ individuals from the donor population; the a_j's dictate the subpool configuration. Blood sera is collected from all n_1 individuals and is indexed in such a way that the LOD reading for every sample is denoted by $\{Y(i_1, \ldots, i_N), 1 \leq i_1 \leq a_1, \ldots, 1 \leq i_N \leq a_N\}$. The individuals are tested and classified according to the following multistage screening procedure (see Figure 7 for a simple example): Start by obtaining Y_0, the LOD reading of the pool composed of all n_1 individuals. Based on Y_0, decide whether all individuals in the pool can be classified as HIV negative or HIV positive; if so, stop testing. If not, then subdivide the population into a_1 subpopulations of size $n_2 = \prod_{j=2}^{N} a_j$, with the first subpopulation consisting of all individuals with $i_1 = 1$, the second with $i_1 = 2$ and so on. Obtain the LOD readings $Y_1(1), \ldots, Y_1(a_1)$ for all subpopulations. For each subpopulation, decide whether all individuals should be

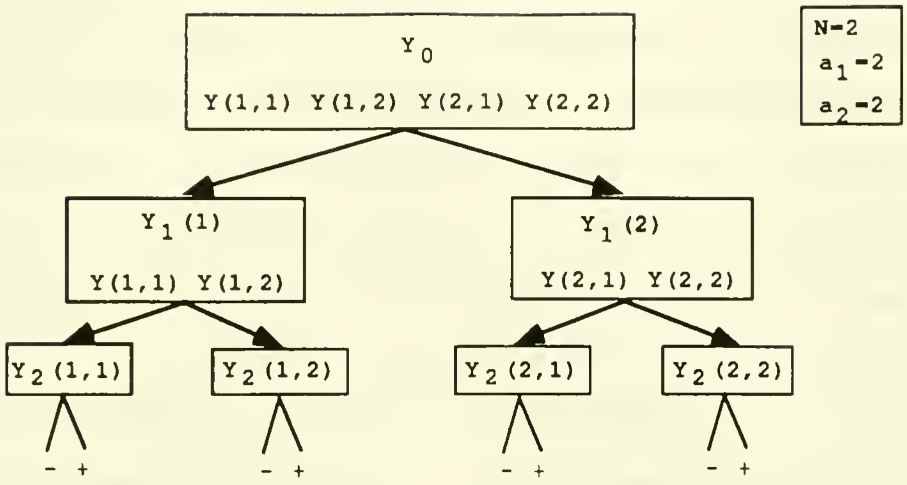


Figure 7: A simple example of a multistage group testing procedure.

classified as HIV negative or positive based on the pair $(Y_0, Y_1(j))$; if so, stop testing. If not, subdivide those subpopulations that require further testing into a_2 subpopulations of size $n_3 = \prod_{j=3}^{N} a_j$. Continue in this vein until either all pools are deemed HIV negative or positive, or stage N, where individual testing is used, is reached. The testing scheme can be equivalently described by a rooted tree, where the nodes of the tree correspond to the different subgroups formed during the procedure.

According to the simplified dilution model (24), the LOD readings are inductively defined by $Y_N(i_1, \ldots, i_N) = Y(i_1, \ldots, i_N)$ and

$$Y_{j-1}(i_1, \ldots, i_{j-1}) = \frac{\sum_{i_j=1}^{a_j} Y_j(i_1, \ldots, i_j)}{a_j}. \tag{31}$$

The state of the system at any stage of the screening process can be described by the LOD readings obtained thus far. If the pool that is currently being screened is composed of all individuals with the first j indices given by $i_1, \ldots, i_j$, then the state is given by $(Y_0, Y_1(i_1), \ldots, Y_j(i_1, \ldots, i_j))$. To simplify notation, we shorten $Y_j(i_1, \ldots, i_j)$ to Y_j for $j = 1, \ldots, N$, and denote the state of the system by $S_j = (Y_0, \ldots, Y_j)$ for $j = 0, \ldots, N$. If the current state of the system is S_j and $j \leq N - 2$, then three decisions are possible: Either declare all individuals in the pool as negative and stop testing, declare all individuals in the pool as positive and stop testing, or subdivide the pool into a_{j+1} subpools of size n_{j+2} and continue testing. Under the first decision, a false negative cost c_{FN} is incurred for each HIV positive individual in the pool, and under the second decision, a false positive cost c_{FP} is incurred for each HIV negative individual in the pool. Under the third decision, a testing cost $c(n_{j+2})$ is incurred for each of the a_{j+1} subpools. By defining $a_{N+1} = 1$, we can adopt the same notation for the decision at stage $N-1$. At stage N (individual testing), the individuals are classified as HIV positive or negative, and the false positive cost c_{FP} or the false negative cost c_{FN} is incurred for any individuals that are misclassified.

If we let $J_j(S_j)$ denote the optimal cost for stages j through N when in state S_j at stage j, then the dynamic programming algorithm is defined inductively by

$$\begin{aligned} J_N(S_N) &= \min\left\{c_{FP}P(Y_N \in P_-|S_N), c_{FN}P(Y_N \in P_+|S_N)\right\}, \qquad (32) \\ J_j(S_j) &= \min\Big\{a_{j+1}(c(n_{j+2}) + E[J_{j+1}(S_{j+1})|S_j]), \\ &\quad c_{FN}\sum_{i_{j+1}=1}^{a_{j+1}} \cdots \sum_{i_N=1}^{a_N} P\left(Y_N(i_1,\ldots,i_N) \in P_+|S_j\right), \\ &\quad c_{FP}\sum_{i_{j+1}=1}^{a_{j+1}} \cdots \sum_{i_N=1}^{a_N} P\left(Y_N(i_1,\ldots,i_N) \in P_-|S_j\right)\Big\} \qquad (33) \end{aligned}$$

for $j = 0, \ldots, N-1$. Because the individual LOD readings of each sample in a pool are iid random variables, equation (33) can be simplified to

$$\begin{aligned} &J_j(S_j) = \min\Big\{a_{j+1}(c(n_{j+2}) + E[J_{j+1}(S_{j+1})|S_j]), \\ &\quad n_{j+1}c_{FN}P(\tilde{Y}_j \in P_+|S_j), n_{j+1}c_{FP}P(\tilde{Y}_j \in P_-|S_j)\Big\} \quad \text{for} \quad j = 0, \ldots, N-1, \qquad (34) \end{aligned}$$

where $\tilde{Y}_j$ is a random variable denoting the individual LOD reading of a generic member of the pool at stage j.

Since the state of the system at stage j is given by the LOD readings obtained through stage j, the dimensionality of the state space grows as the dynamic programming algorithm proceeds; hence, the algorithm in (32) and (34) cannot be efficiently used for numerical calculations. The following proposition shows that we need only keep track of the most recent LOD reading.

Proposition 5 *The state of the system at every stage can be adequately described by the latest LOD reading.*

We prove this proposition using Corollary 2 of Arnold, which is stated here for completeness.

Corollary 2 (Arnold, 1977) *The conditional distribution of Y_{j+1} given S_j is the same as the conditional distribution of Y_{j+1} given Y_j.*

The following lemma, whose proof is given in the Appendix, is also needed:

Lemma 1 *There exists a version of the conditional probability $P(\tilde{Y}_j \in P_+|S_j)$ that depends on S_j only through Y_j; the same is true for $P(\tilde{Y}_j \in P_-|S_j)$.*

Proof of Proposition 5. The proposition can be proved by induction on the dynamic programming algorithm. The proposition is true for $j = N$; assume that it is true for $j = k$, so that $J_k(S_k) = J_k(Y_k)$. By (34) and Lemma 1, J_{k-1} is a function of $J_k(Y_k)$, $P(\tilde{Y}_k \in P_+|Y_k)$ and $P(\tilde{Y}_k \in P_-|Y_k)$; hence, the proposition follows from Corollary 2. ■

Therefore, we can replace the state S_j by the latest LOD reading Y_j, and the optimal decision rule can be described by a set of critical regions R_j^+ and R_j^- for $j = 0, \ldots, N$ such that if $Y_j \in R_j^-$ then all samples in the pool are classified as HIV negative and released for transfusion, if $Y_j \in R_j^+$ then all samples are classified as HIV positive and discarded, and otherwise additional tests are carried out. The critical regions are defined by

$$R_N^+ = \left\{ Y : \frac{f_-(Y)}{f_+(Y)} < \frac{c_{FN}\pi}{c_{FP}(1-\pi)} \right\}, \tag{35}$$

$$R_N^- = \left\{ Y : \frac{f_-(Y)}{f_+(Y)} \geq \frac{c_{FN}\pi}{c_{FP}(1-\pi)} \right\}, \tag{36}$$

$$R_j^+ = \Big\{ Y_j : c_{FP} n_{j+2} P\left(\tilde{Y}_j \in P_-|Y_j\right) \leq$$

$$\min\left\{c_{FN}n_{j+2}P\left(\tilde{Y}_j \in P_+|Y_j\right), c\left(n_{j+2}\right) + E\left[J_{j+1}\left(Y_{j+1}\right)|Y_j\right]\right\}\Big\}, \tag{37}$$

$$R_j^- = \Big\{Y_j : c_{FN}n_{j+2}P\left(\tilde{Y}_j \in P_+|Y_j\right) \leq$$

$$\min\left\{c_{FP}n_{j+2}P\left(\tilde{Y}_j \in P_-|Y_j\right), c\left(n_{j+2}\right) + E\left[J_{j+1}\left(Y_{j+1}\right)|Y_j\right]\right\}\Big\}, \tag{38}$$

where f_- and f_+ denote the normal densities for the HIV negative and HIV positive populations, respectively. Notice that the critical region R_N^- maximizes the power for a simple hypothesis test. Therefore, by the Neyman-Pearson lemma, the proposed classification policy at the individual testing stage not only minimizes the cost for the particular choices of c_{FN} and c_{FP}, it also minimizes the type II error (false positive) for a fixed level of type I error (false negative).

5.2. Structural Properties of the Optimal Policy

Intuitively, one might expect that the optimal classification policy could be characterized by a set of constants $\{c_j^-, c_j^+ : 0 \leq j \leq N\}$ (where $c_N^- = c_N^+$) such that $R_j^- = \left\{Y_j : Y_j \leq c_j^-\right\}$ and $R_j^+ = \left\{Y_j : Y_j \geq c_j^+\right\}$. Such a classification policy for a generalized group testing procedure will be called a *cutoff policy*. Arnold obtained sufficient conditions ensuring the optimality of a cutoff policy for a simpler group testing problem that possesses only two possible classifications. Here, we extend his results to the model in Subsection 5.1.

The following monotonicity notion was introduced in Arnold: The density $g_j(y_j)$ of the LOD reading Y_j has the Mon(j) property if for all nonincreasing functions $h(y)$, the conditional expectation $E(h(Y_j)|Y_{j-1} = s)$ is monotone nonincreasing in s. The following proposition, which is proved in the Appendix, provides sufficient conditions for the optimality of the cutoff policy.

Proposition 6 *A cutoff policy is optimal if the likelihood ratio $\frac{f_+}{f_-}$ is monotone nondecreasing and the density $g_j(y)$ of Y_j has the Mon(j) property for all j.*

The definition of Mon(j) cannot be used for testing whether a density has the required property. Instead, the following proposition can be employed (see the Appendix for a proof).

Proposition 7 *The density $g_j(y)$ has the Mon(j) property if for all j and all y, $P(Y_j \leq y|Y_{j-1} = s)$ is a nonincreasing function of s.*

It turns out that neither of the sufficient conditions in Proposition 6 are satisfied by our data. Recall that f_- is the normal density with mean $\mu_- = -4.82$ and $\sigma_- = 0.42$, and f_+ is normal with mean $\mu_+ = 0.8$ and $\sigma_+ = 1.08$. Because the variation in P_+ is larger than the variation in P_-, $\frac{f_-}{f_+}$ is not monotonically nonincreasing. The individual LOD readings Y_N are distributed as a mixture of two normals. By (31), Y_{N-1} is the average of a collection of iid random variables, each of which is a mixture of normals. It can be shown that the distribution of $Y_N|Y_{N-1}$ is a more complex mixture of normals that does not satisfy the Mon(j) property for our parameter values. Although neither condition in Proposition 6 is satisfied by our data, the optimal cutoff policy performed nearly as well as the overall optimal policy in the computational study described in the next section.

5.3. The Dorfman Policy

In the Dorfman procedure, a pool of a specified size is tested, after which either every sample in the pool is deemed HIV negative, or every sample in the pool undergoes individual testing. Due to its simplicity and effectiveness, this procedure is frequently used in practice for mass screening programs. In particular, recent field studies (e.g., Behets et al., Emmanuel et al. and Kline et al.) in developing countries demonstrate that such procedures can be used to reduce the cost of HIV screening. The complexity of general group testing strategies, such as the one described in Subsection 5.1, renders them more vulnerable to human error. Therefore, the improvement achieved by the more complex testing strategy could be offset by the human errors incurred during implementation.

Using the dynamic program of Subsection 5.1, we can obtain the optimal decision rule for a Dorfman procedure with pool size n by setting $N = 1$ and $n_1 = a_1 = n$, and disallowing the option of discarding a pool that contains more than one sample. However, numerically solving the dynamic programming algorithm requires a discretization of the state space of LOD readings, and can be cumbersome and computationally intensive. Therefore, we propose a relatively simple method for obtaining a near optimal Dorfman policy. Our method relies on two simplifying assumptions: (i) a cutoff policy is employed (that is, the individuals in the pool are transfused if the LOD reading of the pool is below a certain threshold, and each sample in the pool is individually tested otherwise; in the latter case, a second threshold is used to classify individuals as HIV positive or negative), and (ii) the outcome of the pooled test is used to calculate the posterior seroprevalence, but not the posterior conditional densities. The first assumption is not very restrictive, particularly since cutoff policies are the only policies that are apt to be adopted in practice. However, the second assumption is clearly not making the most efficient use of the pooled OD reading.

Consider a Dorfman policy of pool size n applied in a seroprevalence π population. Let Y_i be the LOD reading of the i^{th} individual and Y^p be the LOD reading of the pool of size n. Suppose that x is the cutoff employed for individual testing and z is the cutoff for group testing. If f_- and f_+ are the probability densities for the LOD readings for P_- and P_+ respectively, then the cost of an individual test is

$$C_i(\pi, x) = c(1) + c_{FP}(1-\pi)\int_x^{+\infty} f_-(y)dy + c_{FN}\pi\int_{-\infty}^{x} f_+(y)dy. \tag{39}$$

Let A_{kn} be the event that k out of the n individuals are HIV positive. The cost incurred at the group testing stage of the process is

$$C_g(z) = c(n) + c_{FN}\sum_{k=1}^{n} P(A_{kn})P(Y^p < z|A_{kn})k, \tag{40}$$

where the first term is the testing cost and the second is the misclassification cost of false negatives. Since $Y_1, \ldots, Y_n$ are iid, it follows that $P(Y_i \in P_+|A_{kn}) = \frac{k}{n}$. Under our second assumption, the cost incurred at the second stage of the testing procedure is

$$C_{ig}(z,x) = \sum_{k=0}^{n} P(Y^p > z|A_{kn})P(A_{kn})nC_i(\frac{k}{n},x); \tag{41}$$

hence, the cost per individual is $C(n,x,z) = \frac{1}{n}[C_g(z) + C_{ig}(z,x)]$, or

$$\begin{aligned} C(n,x,z) &= c(n) + c_{FN}\sum_{k=1}^{n}\frac{n!}{k!(n-k)!}\pi^k(1-\pi)^{n-k}kP(Y^p < z|A_{kn}) \\ &+ \sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\pi^k(1-\pi)^{n-k}P(Y^p > z|A_{kn})nC_i(\frac{k}{n},x). \end{aligned} \tag{42}$$

The proposed Dorfman procedure is given by the solution to $\min_{n,x,z} C(n,x,z)$, which can be solved in two stages: Obtain the optimal cutoffs x and z for every n, and then search among the integers for the optimal group size n.

Under our probabilistic assumptions, the LOD reading of a pool composed of k HIV positive and $n-k$ HIV negative individuals is normally distributed with mean $\mu_{kn} = \frac{k\mu_+ + (n-k)\mu_-}{n}$ and variance $\sigma^2_{kn} = \frac{k\sigma^2_+ + (n-k)\sigma^2_-}{n^2}$; that is, $Y_0|A_{kn} \sim N(\mu_{kn}, \sigma^2_{kn})$. The optimal cutoff points are obtained by solving the first and second order optimality conditions. Thus, a locally optimal solution is obtained, which turned out to be globally optimal in our numerical studies.

6. Computational Results

In this section, we assess the relative performance of four testing policies: individual testing (with optimal cutoff values derived from equations (35)-(36)), the heuristic Dorfman policy developed in Subsection 5.3, the optimal Dorfman policy derived from the dynamic programming algorithm and the optimal generalized group testing policy. A wide range of scenarios are considered by varying the seroprevalence and false negative cost. The derivation of the optimal policies is described in Subsection 6.1 and the Monte Carlo simulation model is specified in Subsection 6.2. The policies are tested on the simulation model in Subsection 6.3. In Subsection 6.4, we apply our model to the data from N'tita et al. (1991).

6.1. Computer Implementation

The heuristic Dorfman policy in Subsection 5.2 was implemented using Maple on a Sun Sparc station 20. The partial derivatives of $C(n, x, z)$ with respect to x and z were obtained using symbolic differentiation, and the stationary points were identified using the built-in routine *solve*. Only stationary points lying in the rectangle $[\mu_-, \mu_+] \times [\mu_-, \mu_+]$ were considered, and these points satisfied both the first and second order optimality conditions in all cases. A search over the integers was then employed to obtain the optimal group size n. The optimal group size was found to be bounded above by 20 for $\pi \geq 0.001$ and $c_{FN} \geq 100$; hence, the search was restricted to this region and a procedure similar to interval halving was used.

The implementation of the dynamic programming algorithm is more complex. The continuous state space must be truncated and discretized: our state space consisted of 200 equally spaced points in $[\mu_- - 6\sigma_-, \mu_+ + 6\sigma_+]$, with step size 0.074. Simpson's numerical integration rule with fixed interval size was employed to achieve four digit accuracy.

6.2. The Simulation Model

The analytical model in Section 5 assumes that LOD readings are normally distributed and employs the simplified pooling model (24); however, we believe that this model is not sufficiently realistic to provide a reliable assessment of the policies. Therefore, we resort to Monte Carlo simulation to obtain a more realistic model. However, building a simulation model for this problem is a nontrivial task: There are two possible sources of uncertainty in a pooled LOD reading, the variability in the individual antibody concentrations and the randomness in the manner in which antibodies are detected by ELISAs, and it is difficult to assess the relative impact of each source. Moreover, the pooling GLM (21), which predicts the normalized OD level of a pool as a function of the antibody concentrations of the individuals comprising the pool, cannot be directly simulated because the underlying antibody concentrations are unobservable.

Consequently, we test the policies on two simulation models of varying complexity. The simpler simulation model randomly generates LOD readings for positive and negative individuals from the empirical distributions of Dax that appear in Figure 2(b) and uses the deterministic pooling model (22). Taking $n = 1$ in equation (22) implies that an individual sample's antibody concentration ρ_i is related to its normalized OD level X_i by $\rho_i = \frac{kX_i}{1-X_i}$. Substituting $\frac{kX_i}{1-X_i}$ for ρ_i in (22) gives

$$\ln\left(\frac{X}{1-X}\right) = \ln\left(\frac{\frac{X_1}{1-X_1} + \ldots + \frac{X_n}{1-X_n}}{n}\right), \tag{43}$$

and hence the value of the parameter k need not be estimated for the simulation model. This simulation model is more realistic than the analytical model in two ways: the pooling model (22) does not employ the linear approximation embedded in model (24), and the LOD readings are drawn from the empirical distributions rather than the normal distributions. Although the simulation model ignores the stochastic component of the GLM arising from the binding mechanism of ELISA, both the variability in the antibody concentration and the uncertainty due to the binding mechanism are embedded in the empirical LOD distributions of Dax; hence, the simulation model indirectly captures the second source of uncertainty.

The more complex simulation model attributes the variability of the LOD readings to the variability of antibody concentrations and the stochastic component of the GLM, and is derived from the additional assumptions that (a) the stochastic component of the GLM is negligible for HIV positive individuals and (b) the antibody concentration in HIV negative individuals is deterministic. The first assumption is motivated by the observation appearing near the end of Subsection 4.1 that the normalized OD reading for a HIV positive individual with a given antibody concentration has a very small coefficient of variation. To justify the second assumption, we recall that the normalized OD reading for a HIV negative individual with a given antibody concentration has a coefficient of variation roughly equal to one. The normalized OD readings for HIV negative individuals in Figure 2(a) have mean 0.0083 and standard deviation 0.0085, and hence coefficient of variation 1.016. Therefore, the variability of the normalized OD readings of HIV negative individuals is mostly due to the uncertainty in the binding mechanism that is captured in stochastic component of the GLM, and consequently the variance of the antibody concentration of HIV negative individuals can be approximated by zero.

Let ρ_- denote the deterministic antibody concentration of the HIV negative individuals. To estimate ρ_- from the data, notice that equations (10) and (11) (with the logit function replacing the cloglog function in (11)) imply that the normalized OD readings for HIV negative individuals are $N\left(\frac{\rho_-}{k+\rho_-}, \frac{\rho_-}{(k+\rho_-)^2}\right)$. By setting the mean and variance of this normal distribution equal to the respective mean and variance of the empirical distribution in Figure 2(a), we obtain two equations and two unknowns. The solution to these equations is $\rho_- = 0.968$ and $k = 115.26$. The large discrepancy between the latter value and the estimated value of about 20 from Table 1 may be due to the fact that one estimate is obtained from individual testing data and the other is obtained from a dilution study. To generate the random antibody concentrations ρ_i for HIV positive individuals, we randomly sample normalized OD readings X_i from the empirical distribution in Figure 2(a), and then invert equation (20) to obtain $\rho_i = kX_i/(1 - X_i) = 115.26X_i/(1 - X_i)$. Then we use equations (10) and (21) to calculate the normalized OD reading X for a pool of size n, where the antibody concentrations of the n samples are generated from the seroprevalence and the two distributions specified earlier.

It is not clear to us which of the two simulation models is more realistic; although the more complex model incorporates the stochasticity in the GLM, its use of the normal distribution may lead one to favor the simpler model. It is reassuring to report that the simulation results for the two models are qualitatively nearly identical and quantitatively very similar (expected total costs are within 5% of each other). Hence, in the next subsection, we only report the results for the simple simulation model, and then briefly comment on the results for the complex simulation model.

Now we describe the model parameters. The detailed cost estimates contained in the field study of Behets et al. are employed. They estimated the material and labor cost of testing a single sample to be \$2.12, and the cost of testing a pool containing $n \geq 2$ samples to be $\$2.87 + \$0.083n$. Without loss of generality, we normalize these costs so that the cost of testing a single sample is $c(1) = 1$, and the cost of testing a pool of size n is $c(n) = 1.35 + 0.04n$ for $n \geq 2$. The false positive cost c_{FP} and the false negative cost c_{FN} cannot be as easily estimated. To get a rough estimate for c_{FP}, we note that under the current Red Cross screening protocol, individuals that are found to be HIV positive during an initial ELISA must undergo two additional ELISAs. If at least one of the additional tests is positive, then a highly specific test (Western Blot) is used to verify the individual's serological status. Since ELISA's specificity is more than 0.99, the probability that a noninfected individual with a positive initial ELISA test requires a Western Blot test is approximately (assuming successive ELISA results are independent) $1 - 0.99^2$. The Western Blot test is approximately ten times as costly in materials and labor than an ELISA test. Hence, the expected false positive cost under the Red Cross protocol is $2 + 10(1 - 0.99^2) = 2.199$. This cost may underestimate the true false positive cost because successive ELISA results are not likely to be independent, and the Western Blot test may not be available in developing countries; in the latter case, a human cost may be incurred, particularly if test results are reported to individuals. Hence, we have chosen the conservative estimate of $c_{FP} = 5$.

Since a false negative cost will contaminate the blood supply, these costs are much larger than false positive costs and are very difficult to quantify. Therefore, we consider four different values for c_{FN} (100, 1000, 5000 and 10,000), and combined them with seven different values for the seroprevalence π (ranging from 0.001 to 0.15) to generate 28 different scenarios that span a broad range of possible settings.

For each scenario of the simple simulation model, we randomly generated sample LOD readings using the seroprevalence π and the normal distributions specified earlier. The simulation terminated at the first time after 10,000 simulated pools when the width of the 95% confidence interval for the expected cost dropped below 0.2. To avoid the possibility of sequential dependencies due to any inherent deficiencies of the Turbo Pascal random generator, the *ran0* routine described in Chapter 7 of Press (1988) was used. All four policies were tested on the same random sequence of LOD readings.

6.3. Simulation Results

Dorfman Policies. We begin by comparing the individual testing policy, heuristic Dorfman policy and optimal Dorfman policy; later in this subsection, the generalized group testing policy will be considered. Before assessing the policies' performance, we note that the optimal Dorfman policy turned out to be of the form: Continue at the first stage if the LOD reading is either above a cutoff point or below a second, extremely small, cutoff point. This awkward form arises because under the proposed normal LOD distributions with $\sigma_+ > \sigma_-$, the far left tail of the HIV positive LOD reading eventually dominates the far left tail of the HIV negative reading. However, this phenomenon is due solely to our normality assumption, and does not occur in Figure 2(b). Moreover, such a policy would never be implemented in practice (with good reason), and so we disallowed the option of continuing for extremely low LOD readings, and only report the performance of the Dorfman policy that was optimal within the class of cutoff policies defined in Subsection 5.3. The difference in performance between the overall optimal Dorfman policy and the optimal cutoff Dorfman policy was very small in our numerical study, and hereafter we refer to the optimal cutoff Dorfman policy as the optimal Dorfman policy.

Our main results are reported in Table 2, which describes the policies, and Table 3, which displays their performance in the simulation study. The first column in Table 2 enumerates the 28 scenarios, and the next two columns characterize the scenarios. The final column gives the pool size for each scenario, which was identical for the optimal Dorfman procedure and the heuristic Dorfman procedure. The remaining columns give the LOD cutoff points for the individual testing policy, and for both stages (the first stage is the pooled testing stage and the second stage is the individual testing stage) of both Dorfman procedures. For each scenario, Table 3 gives the 95% confidence interval for the expected total cost of each policy.

The following three observations can be extracted from our numerical study:

(1) The optimal and heuristic Dorfman procedures are quite similar. They both employ identical group sizes for each scenario, and their cutoff points for each stage are relatively close in value in Table 2. Rather surprisingly, as seen in Table 3, the heuristic procedure outperforms the optimal procedure in 23 of the 28 cases, and the expected cost reduction for the heuristic procedure relative to the optimal procedure averages 8.1% over the 28 scenarios. As seen in Table 2, the optimal procedure is slightly more conservative in the choice of cutoff in the pooled testing stage, resulting in policies that are more sensitive, but require more testing. For low seroprevalence, the optimal Dorfman procedure seems to overcompensate for the dilution effect, so that the improved sensitivity does not counteract the resulting increase in monetary testing cost. The strong performance of the heuristic Dorfman procedure is noteworthy, since this policy is much easier to derive than the optimal Dorfman procedure.

(2) Group testing is optimal for all 28 scenarios, and significant savings over individual

			CUTOFF VALUES					
				Heuristic		Optimal		
	π	c_{FN}	Individual	Stage 1	Stage 2	Stage 1	Stage 2	Pool Size
1	0.001	100	-2.99	-4.623	-3.201	-4.653	-2.951	15
2		1000	-3.159	-4.666	-3.328	-4.727	-3.099	12
3		5000	-3.283	-4.728	-3.391	-4.727	-3.247	12
4		10000	-3.338	-4.755	-3.420	-4.801	-3.247	12
5	0.005	100	-3.107	-4.652	-3.228	-4.653	-3.025	13
6		1000	-3.283	-4.633	-3.416	-4.653	-3.247	8
7		5000	-3.414	-4.666	-3.523	-4.727	-3.321	7
8		10000	-3.472	-4.652	-3.586	-4.653	-3.395	6
9	0.01	100	-3.159	-4.623	-3.258	-4.653	-3.099	10
10		1000	-3.339	-4.625	-3.448	-4.653	-3.247	7
11		5000	-3.472	-4.652	-3.569	-4.653	-3.395	6
12		1000	-3.532	-4.626	-3.640	-4.653	-3.513	5
13	0.025	100	-3.231	-4.564	-3.301	-4.579	-3.173	7
14		1000	-3.415	-4.620	-3.485	-4.653	-3.321	6
15		5000	-3.553	-4.638	-3.623	-4.653	-3.513	5
16		10000	-3.615	-4.675	-3.671	-4.727	-3.543	5
17	0.05	100	-3.287	-4.545	-3.324	-4.505	-3.247	5
18		1000	-3.476	-4.593	-3.520	-4.653	-3.395	5
19		5000	-3.618	-4.594	-3.673	-4.653	-3.543	4
20		10000	-3.681	-4.634	-3.727	-4.653	-3.617	4
21	0.1	100	-3.347	-4.514	-3.353	-4.505	-3.321	5
22		1000	-3.54	-4.545	-3.561	-4.579	-3.469	4
23		5000	-3.686	-4.637	-3.694	-4.653	-3.617	4
24		10000	-3.752	-4.679	-3.752	-4.727	-3.691	4
25	0.15	100	-3.384	-4.446	-3.383	-4.431	-3.321	4
26		1000	-3.581	-4.571	-3.573	-4.579	-3.543	4
27		5000	-3.73	-4.665	-3.711	-4.653	-3.691	4
28		10000	-3.797	-4.707	-3.771	-4.727	-3.765	4

Table 2: Optimal Dorfman, heuristic Dorfman, and optimal individual testing policies for the 28 scenarios.

	Expected Total Cost		
	Individual	Heuristic	Optimal
1	1.0069 ± 0.0029	0.2094 ± 0.0026	0.2277 ± 0.0027
2	1.0076 ± 0.0012	0.2843 ± 0.0045	0.4139 ± 0.0061
3	1.0107 ± 0.0014	0.4221 ± 0.0041	0.4245 ± 0.0168
4	1.0118 ± 0.0015	0.5136 ± 0.0044	0.6931 ± 0.0045
5	1.0127 ± 0.0049	0.2998 ± 0.0034	0.2926 ± 0.0034
6	1.0318 ± 0.0278	0.3580 ± 0.0077	0.3771 ± 0.0103
7	1.1035 ± 0.0980	0.4681 ± 0.0487	0.5790 ± 0.0399
8	1.1423 ± 0.0980	0.5314 ± 0.0980	0.5128 ± 0.0925
9	1.0140 ± 0.0046	0.3479 ± 0.0037	0.3734 ± 0.0040
10	1.0528 ± 0.0392	0.4149 ± 0.0131	0.4521 ± 0.0163
11	1.0726 ± 0.0980	0.4735 ± 0.0063	0.5768 ± 0.0980
12	1.0221 ± 0.0021	0.5382 ± 0.0980	0.5960 ± 0.0980
13	1.0147 ± 0.0042	0.4490 ± 0.0044	0.4542 ± 0.0046
14	1.0761 ± 0.0480	0.5325 ± 0.0201	0.5731 ± 0.0239
15	1.3000 ± 0.0980	0.7512 ± 0.0980	0.7608 ± 0.0980
16	1.4672 ± 0.0980	0.9689 ± 0.0980	1.1173 ± 0.0980
17	1.0289 ± 0.0084	0.5964 ± 0.0058	0.5789 ± 0.0055
18	1.1663 ± 0.0759	0.6946 ± 0.0344	0.7783 ± 0.0417
19	1.5831 ± 0.0980	0.9148 ± 0.0980	1.1522 ± 0.0980
20	1.9865 ± 0.0980	1.1504 ± 0.0980	1.4075 ± 0.0980
21	1.0340 ± 0.0095	0.7617 ± 0.0065	0.7649 ± 0.0070
22	1.3203 ± 0.0980	0.9980 ± 0.0652	0.9683 ± 0.0590
23	1.8767 ± 0.0980	1.5287 ± 0.0980	1.5318 ± 0.0980
24	2.3007 ± 0.0980	1.9354 ± 0.0980	1.9686 ± 0.0980
25	1.0389 ± 0.0101	0.9031 ± 0.0078	0.9044 ± 0.0085
26	1.3777 ± 0.0980	1.1965 ± 0.0741	1.1573 ± 0.0688
27	1.9335 ± 0.0980	1.7117 ± 0.0980	1.7876 ± 0.0980
28	2.7271 ± 0.0980	2.5229 ± 0.0980	2.5664 ± 0.0980

Table 3: Simulated performance of the policies for the 28 scenarios.

testing are achieved. The expected cost reduction for the optimal Dorfman procedure relative to individual testing ranges from 5.9% to 77.4%, and the average cost reduction over the 28 scenarios in Table 3 is 39.3%. For the heuristic Dorfman policy, the expected cost reduction relative to individual testing ranges from 7.5% to 79.2% and averages 43.4%. The monetary testing cost is also significantly reduced; the average reduction relative to individual testing is 40% for the optimal policy and 46% for the heuristic policy. Moreover, although we do not show the numbers in Table 3, both Dorfman procedures are highly sensitive and specific. The average sensitivity and specificity over the 28 scenarios are 99.7% and 99.7% for the individual testing policy, 99.8% and 99.7% for the heuristic Dorfman procedure, and 99.8% and 99.7% for the optimal Dorfman procedure. Moreover, the sensitivity of the Dorfman procedures never dropped below 99%.

(3) In Table 2, the optimal cutoff values for individual testing are very similar to the optimal cutoff values for the individual testing stage of the two Dorfman procedures, and range from -3.1 to -3.7; (recall that $\mu_- = -4.82, \sigma_- = 0.42, \mu_+ = 0.80$ and $\sigma_+ = 1.05$). However, the optimal cutoff values for the pooled testing stage of the two Dorfman procedures are much lower, ranging from -4.4 to -4.8. Hence, the Dorfman procedure is able to maintain its high test accuracy by a judicious choice of cutoffs at each stage; more specifically, *the cutoff level is more conservative at the pooled testing stage to compensate for the dilution effect.* In contrast, previous (field and statistical) researchers in pooled testing have assumed that the *the same cutoff level is used at both stages*; in particular, the cutoff level proposed by the test kit manufacturer, which presumably is close to optimal for individual testing, is employed at both stages of the Dorfman procedure.

To assess the performance of the *traditional* Dorfman policy that has been considered in previous studies, we assume that the optimal cutoff level for individual testing (i.e., the fourth column of Table 2) is employed at *both* stages of the procedure. Under this assumption, the optimal value of the pool size n was derived using the cost function (42). The optimal pool size was 15 for scenario 1 and two for the other 27 scenarios. The average expected cost reduction relative to individual testing was 12.78%, which is much smaller than the 39% to 43% reduction achieved by the proposed Dorfman procedures.

To illustrate the predictive power of our model with respect to the traditional Dorfman policy, we consider the study carried out by Behets et al. in Kishasha, Zaire, where the seroprevalence of the 8000 samples was 2.44%. The traditional Dorfman procedure with pools of size ten reduced the monetary cost of HIV screening by 56% relative to individual testing; however, six low reactivity individuals were not detected. We used the Monte Carlo simulation model to calculate the performance of the traditional Dorfman procedure that employed the individual testing cutoff of scenario 15 at both stages. The expected sensitivity of the procedure were 96.4% and the expected monetary testing cost was 0.40 (recall that our testing cost function $c(n)$ is based on the cost model in Behets et al.); hence, our analysis predicts a 60% reduction in monetary testing cost and (0.025)(0.036)(8000)=7.2 false negatives. Therefore, the model accurately captures both the magnitude of the cost savings and the extent of the dilution effect as manifested by the low reactivity individuals that are not detectable in pools. Under the heuristic Dorfman policy for scenario 13, the expected monetary testing cost is 0.43 and the sensitivity is 99.81%, and hence the expected number of false negatives is (0.025)(0.0019)(8000)=0.38. Therefore, we predict that the heuristic Dorfman procedure would not have had any trouble detecting the low reactivity individuals.

Additional scenarios were considered to generate Figure 8, which provides switching curves depicting the optimal group size (as calculated by the heuristic Dorfman procedure) as a function of both the seroprevalence and the false negative cost. As expected, the group size is a decreasing function of both quantities; if the seroprevalence is high, then large group sizes will contain HIV positive individuals with high probability. Similarly, if the false negative cost is high, then smaller group sizes are required to diminish the impact of dilution. Notice that groups of size two or three are never optimal. This phenomenon is due to the pooled testing cost $c(n) = 1.35 + 0.04n$: The cost of constructing the pool is larger than the limited savings realized when the group size is two or three. Finally, in contrast to the case of perfect binary tests, where group testing is optimal if and only if the proportion of defective items is less than $(3 - \sqrt{5})/2 \approx 0.382$ (see Ungar 1960), group testing is only optimal in Figure 8 for significantly lower seroprevalences. The *breakeven* (between individual and group testing) seroprevalence in Figure 8 is a nonincreasing function of c_{FN}, and for $c_{FN} > 100$ it is less than or equal to 0.18. The gap in breakeven seroprevalence between 0.18 and 0.382 is due to the form of our pooled testing cost (the traditional cost is $c(n) = 1$ for all n) and the presence of statistical errors, which becomes increasingly important as seroprevalence increases and leads to a more conservative choice of group size.

The results for the complex simulation model described in Subsection 6.2, although not shown here, are consistent with the results from the simple simulation model. The average sensitivity and specificity over the 28 scenarios are 99.8% and 99.7% for the heuristic, and 99.8% and 99.8% for the optimal. Relative to the individual testing policy, the average expected cost reduction over the 28 scenarios is 42.4% for the optimal Dorfman procedure and 46.1% for the heuristic Dorfman procedure, which are slightly larger than the corresponding values in the simple simulation model. The only qualitatively different result for the more complex simulation model occurs in scenario 4, where the low seroprevalence and high false negative cost lead to a very conservative cutoff at the pooled testing stage for the optimal Dorfman procedure (see Table 2). Consequently, the optimal procedure performed too many individual tests and fared poorly. When the optimal pooled cutoff is increased from -4.801 to -4.76, the monetary testing cost drops drastically, while the sensitivity and specificity remain

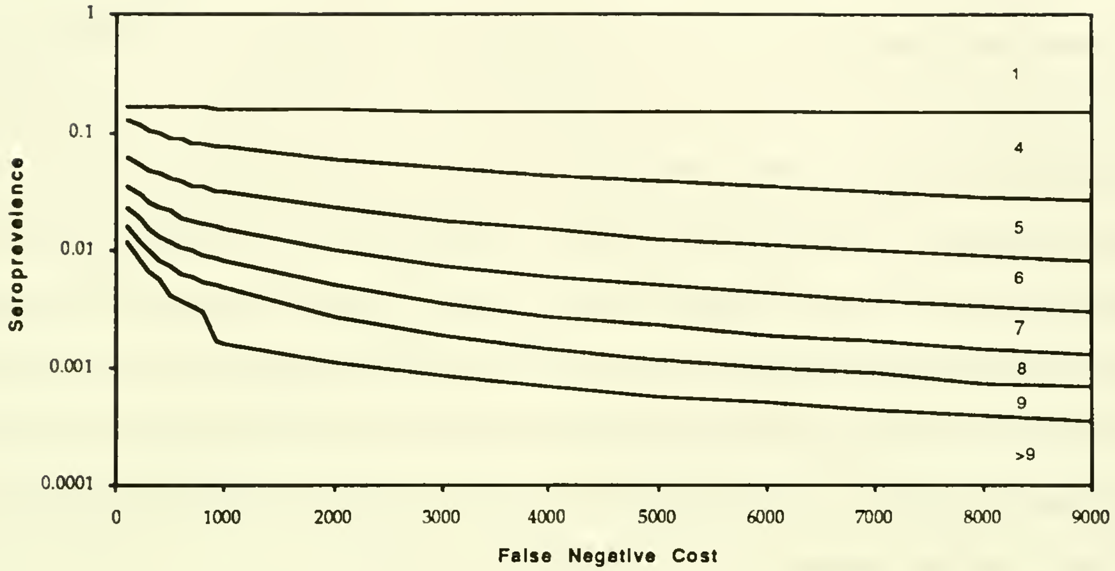


Figure 8: Optimal group size as a function of c_{FN} and π.

unchanged. Hence, the optimal Dorfman procedure is not robust when the seroprevalence is very low and the false negative cost is very high, at least under the complex simulation model.

Generalized Group Testing Policies. We have thus far focused on the Dorfman policy, due to its simplicity and ease of use. However, it is worthwhile comparing the relative performance of the optimal Dorfman policy and a generalized group testing policy. Recall that the dynamic programming algorithm finds the optimal generalized group testing policy for a given initial pool size and subpool configuration. To solve a different dynamic program for each possible initial pool size and subpool configuration is computationally prohibitive. Instead, we investigate the performance of the generalized group testing policy for two scenarios. For scenario 19, where the optimal Dorfman pool size is four, we let the initial pool size $n_1 = 4$, and consider the subpool configuration in Figure 7. For scenario 6, where the optimal Dorfman pool size is 8, we let the initial pool size $n_1 = 8$, and consider the subpool configurations (a) $N = 2$, $a_1 = 2$ and $a_2 = 4$, and (b) $N = 3$ and $a_1 = a_2 = a_3 = 2$. For scenario 19, the expected total cost of the generalized multistage procedure was 1.285 ± 0.0098, which is higher then the cost of either of the proposed Dorfman policies. For scenario 6, subpool configuration (a) outperforms configuration (b), and yields an expected total cost of 0.3558 ± 0.014. This corresponds to a 5.6% expected cost reduction relative to the optimal Dorfman procedure, and a 0.6% reduction over the heuristic policy. In summary, although it is possible to obtain generalized multistage policies that outperform the Dorfman procedures, the additional improvement appears to be offset by the difficulty in deriving and implementing these policies.

6.4. Application

The numerical results in Subsection 6.3 (for example, the switching curves in Figure 8) cannot be universally applied for several reasons. Our numerical results depend upon the HIV positive and HIV negative distributions, which may differ across the world, due to seroconversion rates (a larger seroconversion rate may lead to a fatter left tail of the LOD readings for HIV positive individuals) or the particular strain of virus that is prevalent. Also, the testing cost $c(n)$ may depend upon various economic factors that are distinctive to each country. Nevertheless, we will loosely apply our results in a documented setting to obtain a rough estimate of the benefits that are achievable from group testing.

In Kishasha, Zaire, of the 3741 units of blood transfused in February 1990, 1045 (27.9%) were not screened for HIV infection (see N'tita et al.). Assuming that this was a consequence of budget constraints, we can propose an alternative strategy that will reallocate funds across the transfusion centers so that every blood donor can be tested for antibodies to HIV. Since 72.1% of the units were individually tested, the monetary testing cost of the currently implemented policy is 0.721. Seroprevalence in Zaire is estimated to be about 2.5% (see Behets et al.). Suppose that the individual testing policy under scenario 15 was employed on 72.1% of the units. Since the expected sensitivity and specificity of this policy are 99.8% and 99.6%, respectively, the expected number of infected units that are transfused is $(0.025)(1045) + (0.025)(0.002)(2696) = 26.25$, and the expected number of false positives is $(0.975)(0.004)(2696) = 10.24$. If we use the heuristic Dorfman policy under scenario 15, then the monetary testing cost is 0.54, the expected sensitivity is 99.8% and the expected specificity is 99.6%. Hence, the expected number of infected units that are transfused is $(0.025)(0.002)(3741) = 0.18$, and the expected number of false positives is $(0.975)(0.004)(3741)=14.6$. In summary, pooled testing in this setting reduces the monetary testing cost by 25% *and* reduces the expected number of infected units transfused from 26 to essentially zero. It is clear that pooled testing, if used properly, can save hundreds of lives worldwide.

7. Concluding Remarks

We have developed and validated a mathematical model that captures the dilution effect that occurs when HIV positive sera are pooled with HIV negative sera. The generalized linear models (13) and (21) develop new insights into the nature of the dilution effect, and avoid the heteroscedasticity problem that has plagued the traditional regression models obtained through a purely empirical approach. These GLMs and the simplified pooling model (24) may be useful for other applications besides group testing, and our general approach may be applicable whenever pooled testing is used to identify a disease or contaminant in a liquid.

Our numerical results suggest that the heuristic Dorfman policy derived in Subsection 5.3 provides a cost-effective, accurate and relatively simple alternative to currently implemented HIV screening protocols. This policy can be used in developing countries to safeguard the integrity of the blood supply, and consequently reduce the spread of the AIDS epidemic. While existing field studies and mathematical analyses assume that the same cutoff point is used to classify the pool at both stages of the Dorfman procedure, our analysis shows that only by selecting a different cutoff point at each stage can we ensure that the sensitivity of the test is not compromised. Finally, HIV testing is also extensively used for seroprevalence estimation; in a companion paper, we show how the pooling model developed and validated here can be employed to derive efficient seroprevalence estimates.

Acknowledgment

We are very grateful to Barbara Cahoon-Young, Elizabeth Dax, Esther de Gourville and Richard George for providing data. We also thank Karla Ballman, Barbara Cahoon-Young, Elizabeth Dax, Richard George, David Heymann, Richard Kline, Eugene Litvak, Sheila Mitchell, Peter Page, Constantia Petrou, Chris Stowell, Hiko Tamashiro, and Guido van der Groen for helpful discussions about various aspects of pooled testing. This research is supported by National Science Foundation grant DDM-9057297 and American Foundation for AIDS Research (AmFAR) grant 02100-15-RG.

Appendix

Proof of Proposition 1. By definition, for $s = 0, 1, 2,$

$$\begin{aligned} V_{i,s+1} &= 2^{s+1}(Y^p_{i,s+1} - \mu) && (44) \\ &= 2^{s+1}(\frac{1}{2}Y^p_{is} + \epsilon_{i,s+1} - \mu) && (45) \\ &= V_{is} + 2^s(2\epsilon_{i,s+1} - \mu) && (46) \end{aligned}$$

$$= V_{is} + \hat{\epsilon}_{i,s+1}, \tag{47}$$

where $\hat{\epsilon}_{i,s+1}$ is a random variable with zero mean. ■

Proof of Proposition 2. Since $E(V_{ij}(x) - V_{i0}(x))^2$ is positive semi- definite quadratic form, the minimum is derived from the first order conditions. We have

$$\begin{aligned} E(V_{is}(x) - V_{i0}(x))^2 &= E[(V_{is} - V_{i0} + (2^s - 1)(\mu - x))^2] \qquad (48) \\ &= E[(V_{is} - V_{i0})^2] + 2E[(V_{is} - V_{i0})(2^s - 1)(\mu - x))] \\ &+ (2^s - 1)^2(\mu - x)^2, \qquad (49) \end{aligned}$$

and therefore

$$\begin{aligned} \frac{d}{dx}E(V_{is}(x) - V_{i0}(x))^2 &= -2(2^s - 1)E[(V_{is} - V_{i0})] + 2(2^s - 1)^2(x - \mu) \qquad (50) \\ &= 2(2^s - 1)^2(x - \mu), \qquad (51) \end{aligned}$$

since $E(V_{is}) = E(V_{i0})$ by Proposition 1. Thus, the minimum is attained at $x = \mu$. ■

Proof of Proposition 3. The result follows by calculating $E(\hat{\mu})$ in equation (29) and using Proposition 1, i.e., $E(V_{is} - V_{i0}) = 0$. ■

Proof of Proposition 4. We want to obtain the set of nonnegative weights w_i minimizing the sample variance of the estimator $\hat{\mu}$. Equation (29) can be reexpressed as

$$\hat{\mu} = \mu - \frac{\sum_{i=1}^{10} \sum_{s=1}^{3} w_s(1 - 2^s)(V_{is} - V_{i0})}{10 \sum_{s=1}^{3}[w_s(1 - 2^s)^2]}. \tag{52}$$

Since different samples are independent, the optimal choice of weights is given by the solution to the following minimization problem:

$$\begin{aligned} \text{minimize} \quad & E[\textstyle\sum_{s=1}^{3} w_s(1 - 2^s)(V_{is} - V_{i0})]^2 \qquad (53) \\ \text{subject to} \quad & \textstyle\sum_{s=1}^{3} w_s(1 - 2^s)^2 = \text{constant} \qquad (54) \\ & w_s \geq 0. \end{aligned}$$

The minimization problem can be simplified by defining $x_s = w_s(1-2^s)$, so that the objective function (53) becomes

$$E[\sum_{s=1}^{3} x_s^2 (V_{is} - V_{i0})^2] + 2E[\sum_{s=1}^{3} \sum_{k=s+1}^{3} x_k x_s (V_{is} - V_{i0})(V_{ik} - V_{i0})]. \tag{55}$$

If $k > s$, then

$$\begin{aligned} E[(V_{is} - V_{i0})(V_{ik} - V_{i0})] &= E\left[E[(V_{is} - V_{i0})(V_{ik} - V_{i0})|V_{is}]\right] \quad &(56)\\ &= E\left[(V_{is} - V_{i0})E[(V_{ik} - V_{i0})|V_{is}]\right. \quad &(57)\\ &= E[(V_{is} - V_{i0})^2] \quad &(58)\\ &= \mathrm{Var}(V_{is}). \quad &(59) \end{aligned}$$

Equation (56) follows from the law of total probability for conditional expectations, and equations (58) and (59) follow from the martingale property of the driftless random walk V_{ij}. Combining equations (55) and (59) gives the following, simplified version of the minimization problem:

$$\begin{aligned} \text{minimize} \quad & \textstyle\sum_{s=1}^{3} x_s^2 \mathrm{Var}(V_{is}) + 2\sum_{s=1}^{3}\Big(x_s \mathrm{Var} V_{is}(\sum_{k=s+1}^{3} x_k)\Big) \quad &(60)\\ \text{subject to} \quad & \textstyle\sum_{s=1}^{3}(2^s - 1)x_s = c \quad &(61)\\ & x_s \geq 0, & \end{aligned}$$

where c is the normalization constant.

The optimal solution is obtained by formulating the Lagrangian

$$L(x, \lambda) = \sum_{s=1}^{3} x_s^2 \mathrm{Var}(V_{is}) + 2\sum_{s=1}^{3}\Big(x_s \mathrm{Var}(V_{is}) \sum_{k=s+1}^{3} x_k\Big) + 2\lambda\Big(c - \sum_{s=1}^{3}(2^s - 1)x_s\Big). \tag{62}$$

The Karush-Kuhn-Tucker optimality conditions state that if there exists λ and nonnegative weights w_s satisfying $\frac{\partial L}{\partial w_s} = 0$ and (61), then w is the optimal solution of the minimization problem. This derivative can be written as

$$\mathrm{Var}(V_{is}) \sum_{k=s}^{3} x_k + \sum_{k=1}^{s-1} x_k \mathrm{Var}(V_{ik}) = \lambda(2^s - 1). \tag{63}$$

By the orthogonality property of martingales (Williams 1991, Chapter 12),

$$\mathrm{Var}(V_{is}) = \mathrm{Var}(V_{ik}) + \mathrm{Var}(V_{is} - V_{ik}) \tag{64}$$

for all $s > k$. Combining equation (63) and (64), the condition $\frac{\partial L}{\partial w_s} = 0$ is reformulated as

$$\mathrm{Var}(V_{is}) \sum_{k=1}^{3} x_k + \sum_{k=1}^{s-1} x_k \mathrm{Var}(V_{is} - V_{ik}) = \lambda(2^s - 1). \tag{65}$$

The Karush-Kuhn-Tucker conditions are more conveniently formulated using matrix notation. Let U be the 3×3 matrix defined by $U_{sk} = \mathrm{Var}(V_{is} - V_{ik}) 1_{(s>k)}$, u be the 3×1 vector defined by $u_s = 2^s - 1$, v be the 3×1 vector given by $v_s = \mathrm{Var}(V_{is})$ and e be the 3×1 unit vector. The optimality conditions can then be expressed as

$$\begin{aligned} (x^T e)v + Ux &= \lambda u && (66) \\ u^T x &= c && (67) \\ x &\geq 0. && (68) \end{aligned}$$

In order to obtain the optimal vector x, the vector quantities U and v should be determined. For the random walk model described in Subsection 4.3,

$$v_s = \sigma^2(2^s - 1). \tag{69}$$

From the first row of equation (66), we obtain

$$\sigma x^T e = \lambda, \tag{70}$$

since the first row of U is the zero vector and $u_1 = 1$. Therefore, by combining equations (69) and (70), equation (66) becomes $Ux = 0$. Hence, $x_1 = x_2 = 0, x_3 = \frac{c}{7}$ and $\lambda = \sigma^2 x_3$ satisfy the optimality conditions. This completes the proof of Proposition of 4. The most efficient estimator is

$$\hat{\mu} = \frac{\sum_{i=1}^{10}(8Y_{iN} - Y_{i0})}{70}. \tag{71}$$

■

Proof of Lemma 1: The proof is by induction. The statement is true when $j = N$; assume that it is true for $j = k$. Using the law of total probability for conditional expectations and the fact that all individuals in the pool are indistinguishable, we can write $E[I_{\{\tilde{Y}_k \in P_+\}}|S_{k-1}] = E\left[E[I_{\{\tilde{Y}_k \in P_+\}}|S_k]|S_{k-1}\right]$, and by the induction hypothesis, the right side is equal to $E\left[E[I_{\{\tilde{Y}_k \in P_+\}}|Y_k]|S_{k-1}\right]$. By Corollary 2, this expression is equivalent to $E\left[E[I_{\{\tilde{Y}_k \in P_+\}}|Y_k]|Y_{k-1}\right]$, which depends only on the latest LOD observation Y_{k-1}. ■

Proof of Proposition 6: The proof is by induction. Recall that from equation (34) and Proposition 5,

$$J_j(Y_j) = \min\Big\{a_{j+1}(c(n_{j+2}) + E[J_{j+1}(Y_{j+1})|Y_j]),$$
$$n_{j+1}c_{FN}P(\tilde{Y}_j \in P_+|Y_j), n_{j+1}c_{FP}P(\tilde{Y}_j \in P_-|Y_j)\Big\} \text{ for } j = 0, \ldots, N-1. \tag{72}$$

Let $\hat{J}_j(Y_j) = J_j(Y_j) - n_{j+1}c_{FN}P(\tilde{Y}_j \in P_+|Y_j)$, and $F(Y_N) = c_{FP}P(\tilde{Y}_N \in P_-|Y_N) - c_{FN}P(\tilde{Y}_N \in P_+|Y_N)$. By the law of total probability and the fact that all individuals in the pool are indistinguishable,

$$\begin{aligned} & a_{j+1}E\left[J_{j+1}(Y_{j+1})|Y_j\right] - n_{j+1}c_{FN}P(\tilde{Y}_j \in P_+|Y_j) \\ = \;& a_{j+1}E\left[J_{j+1}(Y_{j+1}) - n_{j+2}c_{FN}P(\tilde{Y}_{j+1} \in P_+|Y_{j+1})|Y_j\right] \\ = \;& a_{j+1}E\left[\hat{J}_{j+1}(Y_{j+1})|Y_j\right], \end{aligned} \tag{73}$$

and

$$\begin{aligned} & n_{j+1}\left(c_{FP}P(\tilde{Y}_j \in P_-|Y_j) - c_{FN}P(\tilde{Y}_j \in P_+|Y_j)\right) \\ = \;& n_{j+1}E\left[c_{FP}P(\tilde{Y}_N \in P_-|Y_N) - c_{FN}P(\tilde{Y}_N \in P_+|Y_N)|Y_j\right] \\ = \;& n_{j+1}E\left[F(Y_N)|Y_j\right]. \end{aligned} \tag{74}$$

Subtracting $n_{j+1}c_{FN}P(\tilde{Y}_j \in P_+|Y_j)$ from both sides of (71) yields

$$\hat{J}_j(Y_j) = \min\left\{a_{j+1}(c(n_{j+2}) + E\left[\hat{J}_{j+1}(Y_{j+1})|Y_j\right]), 0, n_{j+1}E\left[F(Y_N)|Y_j\right]\right\}. \tag{75}$$

Straightforward algebraic manipulations show that

$$F(Y_N) = \frac{c_{FP}(1-\pi) - c_{FN}\pi\frac{f_+(Y_N)}{f_-(Y_N)}}{(1-\pi) + \pi\frac{f_+(Y_N)}{f_-(Y_N)}}; \tag{76}$$

hence, $F(Y_N)$ is monotone nonincreasing by the assumed monotonicity of $\frac{f_+}{f_-}$. Since $\hat{J}_N(Y_N) = \min\{0, F(Y_N)\}$, the function $\hat{J}_N(Y_N)$ is monotone nonincreasing and the unique root of $F(Y_N)$ gives the optimal cutoff c_N^- for stage N.

To prove inductively that $R_j^- = \{Y_j : Y_j \le c_j^-\}$ for some c_j^-, let us assume that $\hat{J}_{j+1}(Y_{j+1})$ is monotone nonincreasing. Then by the Mon(j) property, the nonzero terms on the right side of (74) are also monotone nonincreasing. Therefore, there exists c_j^- (the minimum of the roots of the two terms) such that $\hat{J}_j(Y_j) = 0$ if and only if $Y_j < c_j^-$; thus, $R_j^- = \{Y_j : Y_j \le c_j^-\}$. Moreover, $J_j(Y_j)$ is monotone nonincreasing. This completes the first part of the proof. Similar arguments establish that $R_j^+ = \{Y_j : Y_j \ge c_j^+\}$. ■

Proof of Proposition 7. It is known that $E(X) = \int_{\Re} P(X \ge x)dx$ and $E(X|Z) = \int_{\Re} P(X \ge x|Z)dx$. For a nondecreasing function h and $z_1 > z_2$,

$$\begin{aligned} & E[h(X_i)|Z = z_1] - E[h(X_i)|Z = z_2] = \\ = & \int_{\Re} P(h(X_i) \ge x|Z = z_1)dx - \int_{\Re} P(h(X_i) \ge x|Z = z_2)dx && (77) \\ = & \int_{\Re} \{P(X_i \le h^{-1}(x)|Z = z_1) - P(X_i \le h^{-1}(x)|Z = z_2)\}dx && (78) \\ \le & \ 0. \end{aligned}$$

■

References:

Arnold, S.F. 1977. Generalized Group Testing. *Annals of Statistics* **5**, 1170-1182.

Behets, F., S. Bertozzi, M. Kasali, M. Kashamuka, L. Atikala, C. Brown, R. W. Ryder and C. Quinn. 1990. Successful use of Pooled Sera to Determine HIV-1 Seroprevalence in Zaire with Development of Cost-Efficiency Models. *AIDS* 4, 737-741.

Burns, K. C. and C. A. Mauro. 1987. Group Testing with Test Error as a Function of Concentration. *Commun. Statist.-Theory Meth.* **16**, 2821-2837.

Cahoon-Young, B., A. Chandler, T. Livermore, J. Gaudino and R. Benjamin. 1989. Sensitivity and Specificity of Pooled Versus Individual Sera in Human Immunodeficiency Virus Antibody Prevalence Study. *J. Clinical Microbiology* **27**, 1893-1895.

Cahoon-Young, B. A. Chandler, T. Livermore, J. Gaudino and R. Benjamin 1992. Optimal Pool Size for Determination of HIV Prevalence in Low Risk Populations. Presented at the HIV/AIDS Surveillance Workshop. South San Fransisco, CA.

Cox, D. R. and D. V. Hinkley. 1974. *Theoretical Statistics.* Chapman and Hall, London.

Dax, E. M. Director, National HIV Reference Laboratory, Melbourne, Australia. 1993. Private Correspondence.

de Gourville, E. Research Associate, CAREC, Trinidad W.I. 1992. Private Correspondence.

Dorfman, R. 1943. The Detection of Defective Members of Large Populations. *Ann. Math. Stat.* **44**, 436-441.

Emmanuel, J. C., M. T. Bassett, H. J. Smith and J. A. Jacobs. 1988. Pooling of Sera for Human Immunodeficiency Virus (HIV) Testing: An Economical Method for use in Developing Countries. *J. Clinical Pathology* **41**, 582-585.

Fisher, R. A. 1922. On the Mathematical Foundations of Theoretical Statistics. *Phil. Trans. R. Soc.* **222**, 309-368.

George, R. J. Chief. Dev. Technology Section, Division HIV/AIDS, Center for Disease Control, Atlanta, GA. 1992. Private Correspondence.

George, R. J. and G. Schochetman. 1985. Serological Tests for the Detection of HIV Infection. In *AIDS Testing, Methodology and Management Issues*, G. Schochetman and J. R. George, (ed.), Springer-Verlag, New York, 49-69.

Hastie, T.J. and D. Pregibon. 1992. Generalized Linear Models. In *Statistical Models in S*, J.M. Chambers and T.J. Hastie, (ed.). Wadsworth & Brooks/Cole Computer Science Series, California, 195-247.

Hull, B. 1991. Serum Pooling for HIV Screening in Trinidad and Tobago. Carribean Epidimology Center Technical Report.

Hwang, F. K. 1976. Group Testing with a Dilution Effect. *Biometrika* **63**, 671-673.

Hwang, F. K. 1984. Robust Group Testing. *J. Quality Technology* **16**, 189-195.

Johnson, N. L., S. Kotz and X. Wu. 1991 *Inspection Errors for Attributes in Quality Control.* Chapman and Hall, London.

Kline, R. L., T. A. Brothers, R. Brookmeyer, S. Zegger and T.C. Quinn. 1989. Evaluation of Human Immunodeficiency Virus Seroprevalence in Population Surveys using Pooled Sera. *J. Clinical Microbiology* **27**, 1449-1452.

Ledro-Monroy, G. and E. Archbold. 1990. HIV Serum Pooling Study. Cruz Roja Ecuatoriana.

Litvak, E., X. M. Tu and M. Pagano. 1992. Screening for the Presence of HIV by Pooling

Sera Samples: Simplified Procedures. Working Paper, Harvard School of Public Health.

Madansky, A. 1988. *Prescriptions for Working Statisticians.* Springer-Verlag, New York.

McCullagh, P. and J. A. Nelder. 1989 *Generalized Linear Models, 2nd Edition.* Chapman and Hall, London.

N'tita, I., K. Mulunga, C. Dulat, D. Lusamba, T. Rehle, R. Korte and H. Jagger. 1991. Risk of Transfusion-Associated HIV Transmission in Kishasha, Zaire. *AIDS* **5**, 437-439.

Press, W. H., B. P. Flanney, S. A. Teukolsky and W. T. Vetterling. 1988. *Numerical Recipes in C, The Art of Scientific Computing.* Cambridge University Press, Cambridge.

Tamashiro, H., W. Maskill, J. Emmanuel, A. Fauquex, P. Sato and D. Heymann. 1993. Reducing the cost of HIV antibody testing. *Lancet* **342**, 87-90.

Thompson, K.H. 1962. Estimation of the proportion of vectors in a natural population of insects. *Biometrics* **18**, 568-578.

Tijssen, P. 1985. *Laboratory Techniques in Biochemistry and Molecular Biology.* Vol. **15**. Elsevier, Amsterdam.

Unger, P. 1960. The cutoff point for group testing. *Communication on Pure and Applied Mathematics* **13**, 49-54.

Williams, D. 1991. *Probability with Martingales.* Cambridge University Press, Cambridge, England.